Tests d'hypothèses en BTSA : épisode 1

Yannick Le Bastard - LEGTA Frédéric Bazille

24 août 2024

Table des matières

Résumé : Ce document inaugure une trilogie destinée aux étudiants en statistiques appliquées (psychologie, pharmacie, BTS) ainsi qu'aux enseignants qui souhaitent se former à la notion de tests d'hypothèse.

Le premier épisode souhaite introduire doucement mais sûrement le lecteur à la notion de tests d'hypothèses, ainsi que détailler a minima les techniques qui leur sont rattachées. Un accent particulier sera porté sur la délicate notion de $p-$valeur si utile en pratique mais parfois mal comprise. L'outil informatique (via les logiciels R et Geogebra) nous servira d'illustration dans divers contextes pratiques. **L'accent est mis sur Geogebra pour des raisons pédagogiques mais R et Python seront mis à l'honneur dans les TP**. De nombreux Quiz émailleront ces trois documents dans le but de familiariser le lecteur avec ces concepts-clefs. Les solutions seront presque toutes données à la fin.

Le second épisode couvrira les tests d'indépendance du khi-deux, les tests de préférence, la régression linéaire, les tests de normalité, l'ANOVA, les tests non paramétriques, etc.

Enfin, le troisième et dernier épisode concernera des cas pratiques d'ACP et d'AFC ou de classification hiérarchique ascendante, utiles dans les classes de BTSA.

1 Principes d'un test d'hypothèses

Nous supposons le lecteur familier avec les notions de **population**, d'**échantillon** aléatoire simple, et d'**estimation**. Les lois usuelles comme la **loi normale**, la **loi de Student** ou la **loi du khi-deux** sont supposées connues.

1.1 Scène et acteurs

Exemple 1 :
Une pandémie sévit dans une certaine région du globe. Un test de dépistage est disponible pour celle-ci, mais n'est pas fiable à 100%. Nous cherchons à déterminer si un individu prélevé au hasard dans la population est sain ou malade à l'aide de ce test.

Il est usuel en statistiques de définir deux situations incompatibles entre lesquelles il nous faudra faire un choix :

1. *L'hypothèse nulle*, notée H_0 : "Il n'y a pas d'effet significatif",
2. *L'hypothèse alternative*, notée H_a ou H_1 : "Il y a un effet significatif"

Mais qu'est-ce qu'un effet non significatif dans le cas présent ? Ou son contraire ?
Il semble naturel ici de reformuler nos deux hypothèses par :

— H_0 : "L'individu est sain"
— H_1 : "L'individu est malade"

Ce que nous pouvons résumer dans le tableau suivant :

Décision \ Réalité	H_0 vraie	H_1 vraie
H_0 décidée	Vrai négatif	Faux négatif
H_1 décidée	Faux positif	Vrai positif

Quiz 1 : Nous supposons que la probabilité pour un individu donné d'être malade est de 0,005. La **sensibilité** (Se) est la probabilité qu'un test réalisé sur une personne malade se révèle positif ; autrement dit, que le test soit positif sachant que la personne est malade. La **spécificité** (Sp) est la probabilité qu'un test réalisé sur une personne saine se révèle négatif ; autrement dit, que le test soit négatif sachant que la personne n'est pas malade. On note VP (resp. VN) la probabilité d'avoir un vrai positif (resp. un vrai négatif) et FP (resp. FN) la probabilité d'avoir un faux positif (resp. un faux négatif).

1. La sensibilité d'un test est égale à :
   (a) $\frac{VP}{VP+VN}$ (b) $\frac{VP}{FP+VN}$ (c) $\frac{VP}{VP+FN}$ (d) $\frac{VP}{FP+FN}$
2. la spécificité d'un test est égale à :
   (a) $\frac{VN}{VP+VN}$ (b) $\frac{VN}{FP+VN}$ (c) $\frac{VN}{VP+FN}$ (d) $\frac{VN}{FP+FN}$
3. La sensibilité du test est estimée à 0,997. La probabilité qu'un individu aléatoire soit testé positif et soit malade est de :
   (a) 0,0997 (b) 0,0499 (c) 0,00499 (d) 0,0055
4. La probabilité qu'un individu aléatoire soit testé négatif et soit sain est de 0,9552. La spécificité du test est de :
   (a) 0,094 (b) 0,96 (c) 0,0095 (d) 0,9755
5. Un laboratoire effectue ce test sur 10 individus indépendants choisis aléatoirement.

(a) La probabilité qu'un individu donné soit déclaré sain à l'issue d'un test est égale à :
   (a) 0,005 (b) 0,0448 (c) 0,995 (d) 0,9552
(b) La probabilité qu'au moins un individu soit déclaré malade à l'issue des 10 tests est égale à :
   (a) 0,0489 (b) 0,3677 (c) 0,6323 (d) 0,1385

1.2 Erreurs de type 1 et de type 2 - Puissance d'un test

La lecture du tableau précédent nous amène à considérer deux types d'erreurs : déclarer un individu sain comme malade (Faux positif) ou déclarer un individu malade comme sain (Faux négatif).
Le cas présent (très artificiel) nous permet de calculer exactement ces erreurs.

Ainsi, en reprenant les données du Quiz 1 :
$\alpha = P_{H_0 \text{ vraie}}(\text{Décider } H_1) = P_{\text{être sain}}(\text{avoir un test positif}) = 1 - \text{Sp} = 0,04.$
$\beta = P_{H_1 \text{ vraie}}(\text{Décider } H_0) = P_{\text{être malade}}(\text{avoir un test négatif}) = 1 - \text{Se} = 0,003.$

Définition 1-2-1 :

1. On appelle *erreur de première espèce* et on note α la probabilité de décider H_1 alors que H_0 est vraie ($i.e$ la probabilité de rejeter H_0 à tort).
2. On appelle *erreur de seconde espèce* et on note β la probabilité de décider H_0 alors que H_1 est vraie ($i.e$ la probabilité d'accepter H_0 à tort).

Le tableau précédent peut se reformuler ainsi :

Réalité / Décision	H_0 vraie	H_1 vraie
H_0 décidée	$1-\alpha$	β
H_1 décidée	α	$1-\beta$

Définition 1-2-2 : On appelle *puissance* d'un test statistique la probabilité de décider H_1 (dire qu'il y a un effet significatif) alors que H_1 est vraie (il y a effectivement un effet significatif). Autrement dit, la puissance du test est égale à $1-\beta$.

Le lecteur se convaincra sans peine que la puissance du test précédent est de 0,997 (c'est la sensibilité Se). Ce qui, nous le verrons en détail, est excellent. Mais nous avions toutes les données en main !
Une autre remarque : si p désigne la proportion de malades dans la population, Se la sensibilité du test et Sp la spécificité du test, la probabilité qu'un individu soit déclaré positif à l'issue du test est égale à :

$$pSe + (1-p)(1-Sp)$$

Un autre exemple éclairant est la tenue d'un procès. À vous de formuler les hypothèses H_0 et H_1 et d'interpréter les deux types d'erreur.

Passons maintenant à un cas plus calculatoire.

Exemple 2 : On s'intéresse à la masse moyenne d'un élevage de lapins dans la ville de Clapiers.
On confronte deux hypothèses :

— H_0 : la masse des lapins est distribuée selon une loi normale de moyenne $\mu_0 = 1,4$ kg et d'écart-type $\sigma_0 = 0,3$ kg,

— H_1 : la masse des lapins est distribuée selon une loi normale de moyenne $\mu_1 = 1,7$ kg et d'écart-type $\sigma_1 = 0,4$ kg.

On prélève un échantillon aléatoire simple de 16 lapins dans cet élevage. Nous rappelons que si $X \hookrightarrow \mathcal{N}(\mu;\sigma)$, alors la variable aléatoire $\bar{X}$ qui à tout E.A.S (échantillon aléatoire simple) de taille n associe la moyenne du caractère observé, suit la loi normale de moyenne μ et d'écart-type $\frac{\sigma}{\sqrt{n}}$ $i.e$ $\bar{X} \hookrightarrow \mathcal{N}\left(\mu;\frac{\sigma}{\sqrt{n}}\right)$.

On se donne la règle de décision suivante : si la masse moyenne des lapins de l'échantillon est inférieure à 1,5 kg, on ne rejette pas H_0, sinon on rejette H_0.

Calculer les risques α et β et en donner une interprétation graphique.

Soit $\bar{X}$ la variable aléatoire qui à tout échantillon aléatoire simple de 16 lapins lui associe sa masse moyenne.

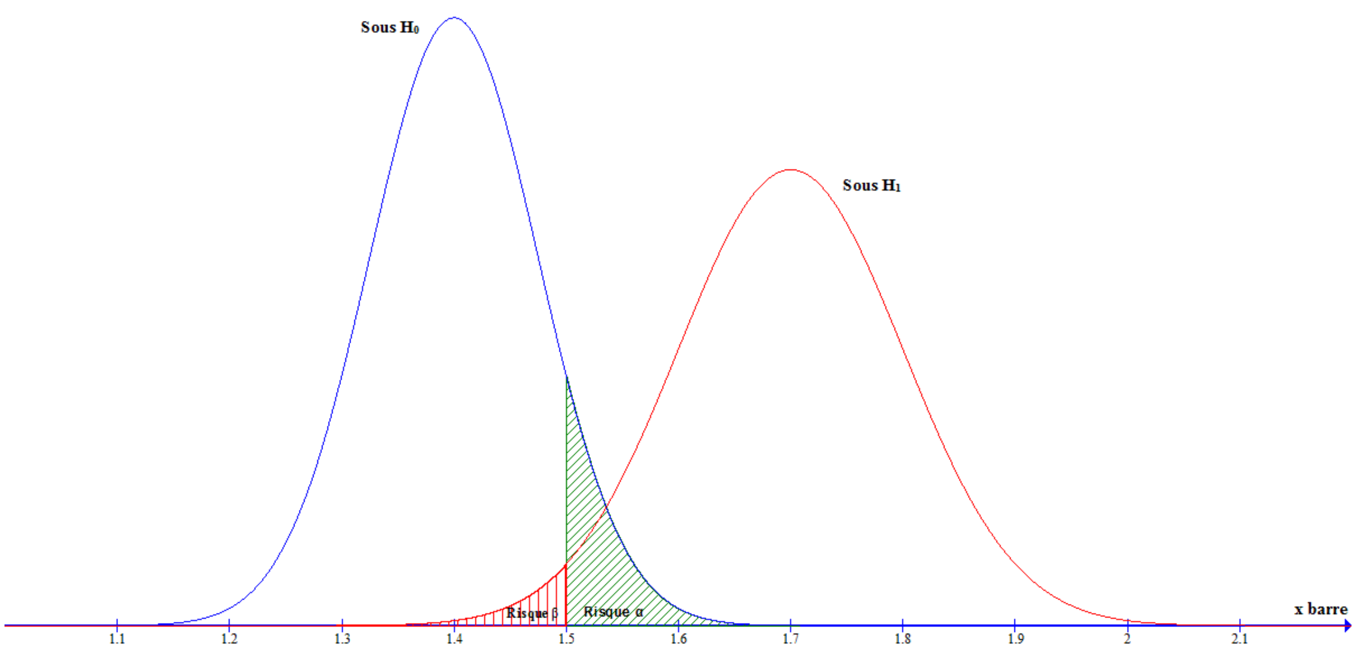
Sous H_0, $\bar{X} \hookrightarrow \mathcal{N}(1,4;0,075)$.

$\alpha = P_{H_0 \text{ vraie}}(\text{Décider } H_1) = P_{H_0 \text{ vraie}}(\bar{X} \geq 1,5) = 0,0912$.

Sous H_1, $\bar{X} \hookrightarrow \mathcal{N}(1,7;0,1)$.

$\beta = P_{H_1 \text{ vraie}}(\text{Décider } H_0) = P_{H_1 \text{ vraie}}(\bar{X} \leq 1,5) = 0,0228$.

Illustrons ceci graphiquement :



Quiz 2 : Deux scientifiques souhaitent comparer la taille des tardigrades d'un glacier. Pour ceci, ils se réfèrent chacun à deux études distinctes. La première étude, de grande ampleur, affirme que les tailles des tardigrades sont distribuées selon une loi normale de moyenne 1,1 mm et d'écart-type 0,012 mm. La seconde prétend que les tailles des tardigrades sont distribuées selon une loi normale de moyenne 1,116 mm et d'écart-type 0,018 mm.

On prélève un échantillon (assimilé à un échantillon aléatoire simple) de 16 tardigrades de ce glacier. On se donne la règle de décision suivante : si la taille moyenne des tardigrades de l'échantillon est inférieure à 1,106 mm, *on ne rejette pas* les données de la première étude ; sinon, on ne rejette pas les données de la seconde étude.

On appelle μ la taille moyenne des tardigrades de ce glacier. On décide pour hypothèse nulle $H_0 \ : \ \mu = 1,1$ et pour hypothèse alternative $H_1 \ : \ \mu = 1,116$.

1. Le risque α est égal à :
   (a) 0,9996 (b) 0,6915 (c) 0,9772 (d) 0,0228
2. La puissance du test est égale à :
   (a) 0,0004 (b) 0,9869 (c) 0,0478 (d) 0,6915
3. Par quelle valeur faut-il remplacer 1,106 pour que le risque α soit égal à 1% ?
   (a) 1,108 (b) 1,109 (c) 1,107 (d) 1,105
4. Augmenter la taille de l'échantillon (plusieurs réponses possibles) :
   (a) augmente le risque α (b) diminue le risque α (c) augmente le risque β (d) diminue le risque β
5. On appelle *courbe de puissance* la courbe représentative de la fonction P définie sur $[0;1]$ à valeurs dans $[0;1]$, qui à chaque risque α associe la puissance $1-\beta$ du test.
   (a) La fonction P est croissante (b) La fonction P est décroissante (c) La fonction P n'est pas monotone

La règle de décision précédente entre nos deux hypothèses H_0 et H_1 parait bien arbitraire. En fait, il est possible, pour un risque α fixé a priori, et pour ce genre d'hypothèses simples ($\mu = \mu_0 \textit{ vs } \mu = \mu_1$), de déterminer la valeur critique[1] qui sépare les zones de non-rejet et de rejet de H_0 avec une puissance maximale. Dans la pratique, les hypothèses H_0 et H_1 ne jouent pas un rôle symétrique. Aussi, nous nous fixerons *a priori* le risque de première espèce α et supposerons l'hypothèse nulle H_0 vraie. Il s'agit donc d'**une hypothèse sur la population étudiée** (condition très forte). Et ce sont les données de notre échantillon qui vont nous permettre de pencher vers une hypothèse plutôt que pour l'autre. L'outil essentiel, qui dépendra de la nature du paramètre de la population que nous souhaitons évaluer, sera une *variable de décision i.e* une variable aléatoire. Nous pouvons donc définir ...

Définition 1-2-3 : Un **test statistique** est un mécanisme qui permet de trancher entre deux hypothèses incompatibles H_0 et H_1 au vu des résultats d'un échantillon, en *quantifiant* le risque associé à la décision prise.

Remarque 1-2-4 :

1. L'hypothèse H_0 joue souvent un rôle particulier. Imaginons par exemple qu'un boulanger prétende que les baguettes de sa production pèsent en moyenne 250 g. Un client suspicieux pèse chaque jour les baguettes qu'il lui achète et se rend compte que sur un échantillon de 30 baguettes le poids moyen est de 249,5 g. Le cas présent nous incite à poser pour hypothèse nulle $H_0 \ : \ \mu = 250$ et pour hypothèse alternative $H_1 \ : \ \mu < 250$, où μ désigne la masse moyenne d'une baguette de la production.
   L'hypothèse nulle vient du fait que l'on peut la réécrire sous la forme $\mu - 250 = 0$; on peut aussi l'interpréter comme "nulle" **ne prouve rien si elle est acceptée**. Quant à son alternative, nous parlerons de *test unilatéral à gauche*.
   L'hypothèse alternative $\mu \neq 250$ (resp. $\mu > 250$) nous conduit à un *test bilatéral* (resp. à un *test unilatéral à droite*).
2. C'est le contexte qui décidera de la nature de l'hypothèse alternative que nous choisirons. Remarquons que dans le cas des tests bilatéraux, l'erreur de première espèce sera "coupée en deux parties d'aires égales". Nous mettrons ceci en lien avec les intervalles de fluctuation.

1. à l'aide du théorème de Neyman et Pearson que nous admettrons ici

Définition 1-2-5 : Une hypothèse est dite **simple** si la loi de la variable aléatoire X en jeu est totalement spécifiée quand cette hypothèse est réalisée. Sinon, on parle d'hypothèse **multiple** ou **composite**.

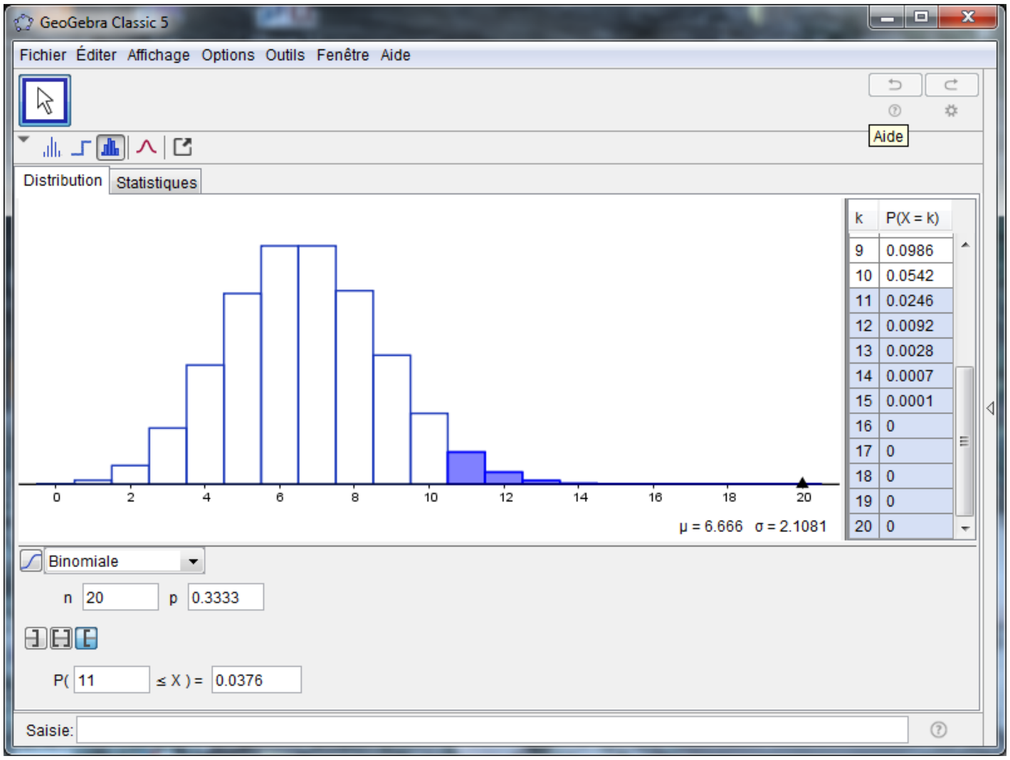
Remarque 1-2-6 :

1. Si θ est un paramètre de la population à estimer, l'hypothèse $\theta = \theta_0$ est une **hypothèse simple** ; l'hypothèse $\theta \neq \theta_0$ est une **hypothèse composite** : θ peut prendre une infinité de valeurs.
2. L'écriture $\theta = \theta_0$ ne signifie pas une égalité mathématique au sens propre du terme mais s'interprète plutôt comme θ et θ_0 ne sont pas significativement différents. **Les éventuelles différences observées sont imputables à la fluctuation d'échantillonnage**.

Exemple 3 : On initie une séance de dégustation à l'aveugle auprès d'un groupe de 20 amateurs de vin. Chacun d'entre eux a trois verres opaques à disposition. Deux verres contiennent le même vin et le troisième contient un vin différent. On admet que la répartition des trois vins dans chaque lot (AAB ou BBA) est effectuée au hasard et que les dégustateurs sont indépendants. Par ailleurs, chacun d'entre eux donne obligatoirement une réponse à la question "Quel verre contient le vin différent ?" [2]. Notre panel nous fournit 12 bonnes réponses.
Soit X la variable aléatoire qui à tout échantillon aléatoire simple de 20 dégustateurs associe le nombre de bonnes réponses obtenues.
Supposons que tous les amateurs répondent au hasard. Le lecteur se convaincra aisément que X suit une loi binomiale de paramètres $n = 20$ et $p = \frac{1}{3}$. Ce sera notre hypothèse nulle $H_0 \ : \ p = 1/3$ (pas d'effet dégustateur).

1. Étape 1 : formulation des hypothèses.
   $$\begin{cases} H_0 : p = 1/3 \\ H_1 : p > 1/3 \text{ (test unilatéral à droite)} \end{cases}$$
   Ce choix s'impose de lui-même pour justifier qu'il y a un "effet dégustateur".
2. Étape 2 : Fixons a priori le risque α à 5%.
3. Étape 3 : Sous H_0, nous utilisons la variable X comme variable de décision.
4. Étape 4 : Trouvons le **seuil critique** qui va séparer la **zone de rejet de** H_0 de sa **zone de non-rejet**, que nous qualifierons un peu abusivement de zone d'acceptation de H_0. Un software nous permet de voir que $P(X \geq 10) \geq 0,05$ et que $P(X \geq 11) < 0,05$.

2. Un tel test est appelé *test triangulaire*. Pour plus de détails *cf* `http://www.numdam.org/article/JSFS_1994__135_3_21_0.pdf`



5. Étape 5 : Énonçons la règle de décision :
   — si x (nombre de bonnes réponses) est inférieur ou égal à 10, on ne rejette pas H_0,
   — si $x > 10$ *i.e* si $x \geq 11$ ici, on rejette H_0.
6. Étape 6 : conclusion.
   Le nombre de bonnes réponses est de 12. Nous sommes donc dans la zone de rejet de H_0. Au risque de première espèce de 5%, nous pouvons supposer (mais ce n'est pas une preuve) qu'il est improbable qu'il n'y ait aucun effet "dégustateur".

La loi de X étant discrète, nous ne pouvons pas trouver une valeur réelle unique x_c telle que $P(X > x_c) = 0,05$. Ceci ne sera bien sûr pas le cas avec des variables aléatoires absolument continues telles la loi normale.

Remarque 1-2-7 : Ces six étapes sont classiques pour mener un test d'hypothèses. Et selon la nature du test, nous serons amenés à utiliser une variable de décision qu'il faudra bien définir, soit de manière exacte ou approchée. **Cette variable de décision est construite sous l'hypothèse que H_0 est vraie.**
D'autre part, remarquez que nous ne nous sommes pas basés sur le calcul de $P(X = 12)$ mais bien sur $P(X \geq 12)$: probabilité d'obtenir 12 bonnes réponses ou plus.
Nous aurions pu également dès l'étape 4 calculer $P(X \geq 12)$. C'est ce que l'on appelle une $p-$**valeur** : "*obtenir 12 bonnes réponses ou plus extrême*". On trouve que $P(X \geq 12) = 0,013 < \alpha = 0,05$. C'est une notion que nous détaillerons plus loin.

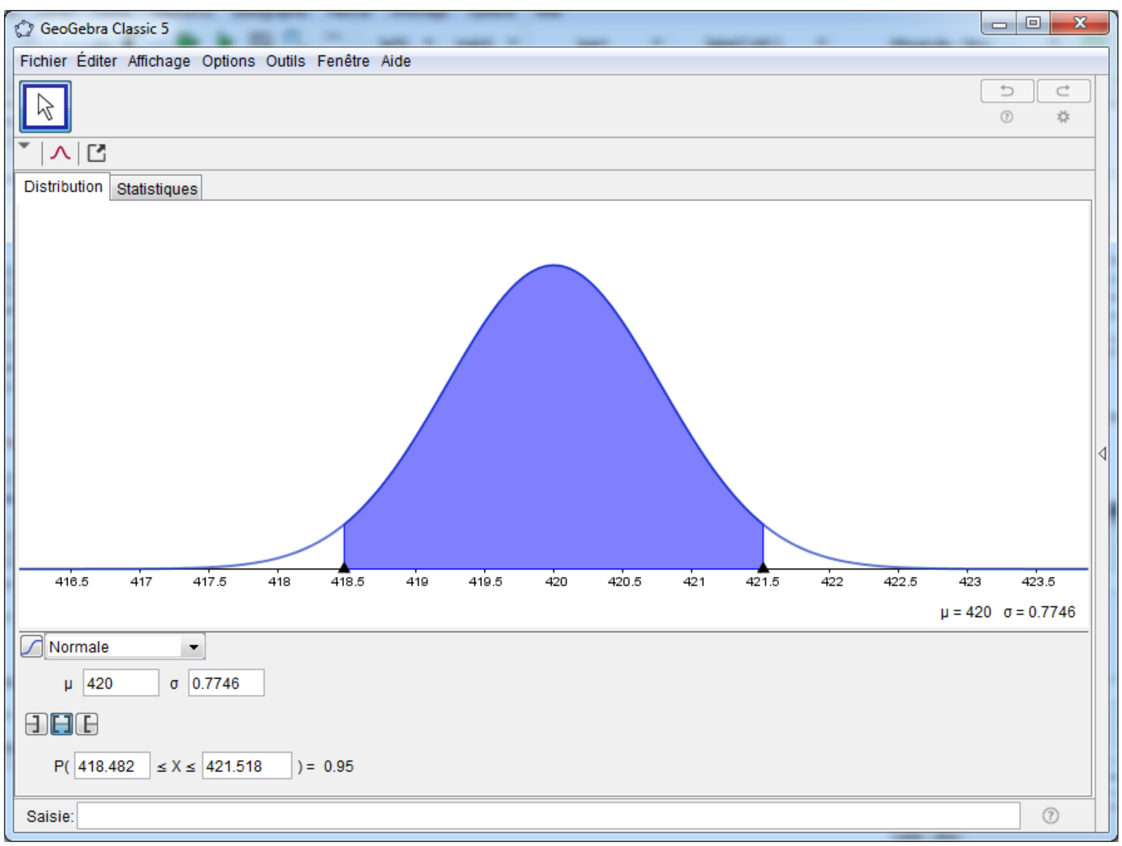
Exercice 1 : Expliquez par exemple la case "nombre de jugements = 11" et "nombre minimum de jugements corrects = 8" pour un seuil de confiance de 99% du document `http://www2.agroparistech.fr/IMG/pdf/Test_triangulaire.pdf`

Exemple 4 : Il est inscrit sur un emballage de boites de conserves une contenance de 420 g. Il est admis que l'écart-type σ de la production est de 3 g et que *la distribution des masses* *est supposée normale.* Lors d'un contrôle qualité, on prélève un échantillon (assimilé à un échantillon aléatoire simple) de 15 boites de conserves. La masse moyenne observée est de 422 g. Peut-on, au risque de première espèce de 5%, supposer que la masse moyenne d'une boite de conserve est :

1. différente de 420 g ?
2. supérieure à 420 g ?

Question 1 : Comme dans l'exemple précédent, nous allons formaliser notre démarche par étapes.

1. Étape 1 : formulation des hypothèses.
   Soit μ la masse moyenne d'une boite de conserve de la production.
   $\begin{cases} H_0 : \mu = 420 \\ H_1 : \mu \neq 420 \text{ (test bilatéral)} \end{cases}$
2. Étape 2 : le risque α est fixé à 5%.
3. Étape 3 : Soit $\bar{X}$ la variable aléatoire qui à tout EAS de 15 boites de conserve associe la masse moyenne observée. Sous H_0, $\bar{X} \hookrightarrow \mathcal{N}\left(420; \frac{3}{\sqrt{15}}\right)$
4. Étape 4 : Trouvons les **seuils critiques** qui vont séparer la **zone de rejet de** H_0 de sa **zone de non-rejet**. Ceci revient à trouver les fractiles a et b d'ordres respectifs 0,025 et 0,975 de $\bar{X}$ *i.e* a est tel que $P(\bar{X} \leq a) = 0,025$ et b est tel que $P(\bar{X} \leq b) = 0,975$. Un software nous permet de voir que $a = 418,482$ et que $b = 421,518$.



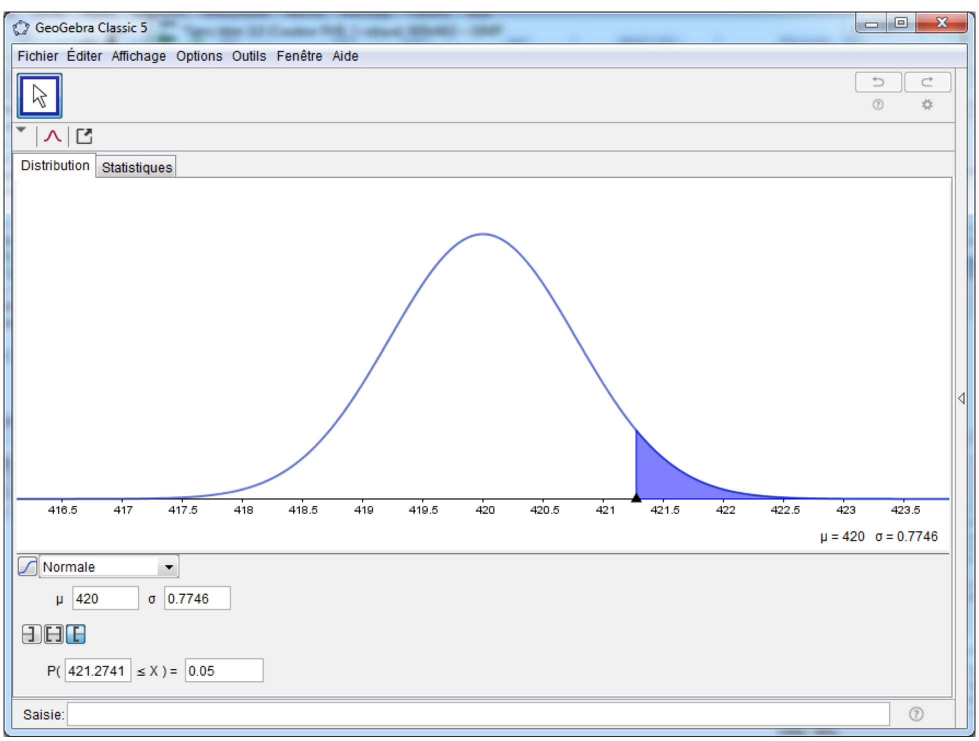
5. Étape 5 : Énonçons la règle de décision :
   — si $\bar{x} \in [418,482; 421,518]$, on ne rejette pas H_0,
   — si $\bar{x} \notin [418,482; 421,518]$, on rejette H_0.

6. Étape 6 : conclusion.
La masse moyenne de l'échantillon est $\bar{x} = 422$. Nous sommes donc dans la zone de rejet de H_0. Au risque de première espèce de 5%, nous pouvons affirmer que $\mu \neq 420$. La machine qui remplit les boites des conserves est certainement déréglée.

Remarque 1-2-8 :

1. $\bar{X} = 420 + \dfrac{3}{\sqrt{15}}U$, où $U \hookrightarrow \mathcal{N}(0;1)$. Nous aurions ainsi pu prendre U comme variable de décision et non $\bar{X}$. Les valeurs critiques sont alors -1,96 et 1,96 et la règle de décision devient : si $u = \dfrac{\bar{x} - 420}{3/\sqrt{15}} \in [-1,96; 1,96]$, on conserve H_0, sinon on rejette H_0. C'est cette option qui est utilisée dans la pratique car les fractiles usuels de la loi normale centrée-réduite sont bien connus et une table de sa fonction de répartition est disponible.
2. Sous $H_0 : \mu = 420$, l'intervalle de non-rejet de H_0 : $[418,482; 421,518]$ apparait comme un ***intervalle de fluctuation (ou de pari)*** de la masse moyenne d'une boite de conserve au seuil de confiance de 95%.
3. Si à partir de l'échantillon, nous avions voulu calculer un ***intervalle de confiance*** de μ, nous aurions trouvé (exercice) : $[420,482; 423,518]$. 420 est en dehors ! Ceci correspond au rejet de l'hypothèse nulle H_0.

Question 2 : la résolution de la question précédente nous amène directement à la même conclusion. Pourquoi ? Car le seuil de rejet de H_0 qui détermine la valeur critique est nécessairement inférieur à 421,518 : notre test est ici un test unilatéral à droite.
La fonction de répartition $F_{\bar{X}}$ de $\bar{X}$ est continue et strictement croissante, donc définit une bijection de $\mathbb{R}$ sur $F_{\bar{X}}(\mathbb{R}) =]0;1[$. Ainsi, $F_{\bar{X}}^{-1}$ définit une bijection strictement croissante de $]0;1[$ sur $\mathbb{R}$. 1 - 0,05 < 1 - 0,05/2, d'où $F_{\bar{X}}^{-1}(0,95) \leq F_{\bar{X}}^{-1}(0,975)$ (**faire un dessin**).
Un software nous donne comme nouvelle valeur critique $F_{\bar{X}}^{-1}(0,95) = 421,274$.



Remarque 1-2-9 : Ce dernier résultat nous permet de constater qu'un test unilatéral est plus puissant qu'un test bilatéral. Nous reviendrons en détail sur la notion de puissance d'un test (comment l'améliorer notamment) dans la section dédiée aux $p-$valeurs et dans la section compléments théoriques et pratiques.

Quiz 3 : Une affection fongique atteint environ 11% des plantes d'une espèce donnée.
Une année pluvieuse et anormalement chaude, on constate que sur un échantillon supposé aléatoire simple de 200 plantes, 28 d'entre elles présentent cette affection. Soit π la proportion de plantes atteintes cette année par cette affection. Nous posons $H_0 : \pi = 0,11$. Soit F la variable aléatoire qui à tout échantillon aléatoire simple de 200 plantes associe la proportion de plantes atteintes par le champignon.

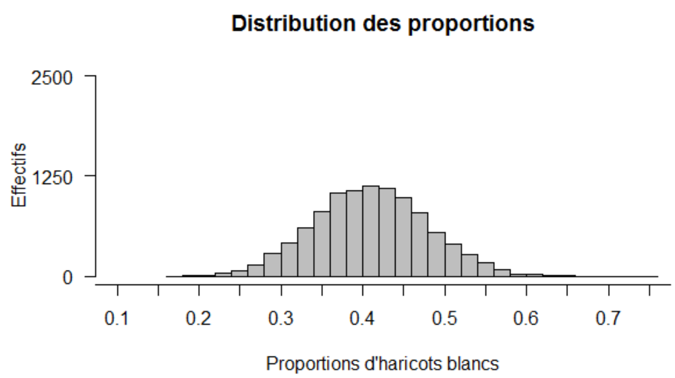
1. La valeur f prise par F dans notre échantillon est de :
   (a) 0,28 (b) 0,11 (c) 0,14 (d) 0,12
2. Au vu de l'échantillon, quelle assertion est-il la plus raisonnable de poser ?
   (a) $H_1 : \pi = 0,11$ (b) $H_1 : \pi > 0,11$ (c) $H_1 : \pi < 0,11$ (d) $H_1 : \pi \neq 0,11$
3. Sous l'hypothèse nulle, $P(F \geq 0,14)$ est égale à :
   (a) 0,11 (b) 0,0876 (c) 0,0749 (d) 0,1094
4. On teste $H_0 : \pi = 0,11$ *vs* $H_1 : \pi > 0,11$ au risque α de 5%. On rejette H_0.
   (a) Vrai (b) Faux
5. sous H_0, le couple $(E(F), \sigma(F))$ est égal à :
   (a) (0,11 ; 19,58) (b) (0,11 ; 4,42) (c) (0,022 ; 0,11) (d) (0,11 ; 0,022)
6. On admet que les conditions pour "approcher" F par une loi normale de paramètres donnés par l'item précédent sont acquises (cette approximation tient au théorème central-limit que nous retrouverons dans la partie Compléments). Si bien que $U = \dfrac{F - E(F)}{\sigma(F)} \hookrightarrow \mathcal{N}(0; 1)$. En prenant U comme variable de décision, la zone de rejet de H_0 est :
   (a) $]1,645; +\infty[$ (b) $]0,12; 0,153[$ (c) $]-\infty; 1,645]$ (d) $]0,1462; +\infty[$
7. Retrouvez-vous la conclusion obtenue avec la loi exacte de F ?
   (a) Oui (b) Non (c) Je ne sais pas

Exemple 5 : Effectuons une petite simulation artificielle avec **R** : le paramètre étudié dans la population, ici une proportion, est supposé inconnu (pas vrai ici). Donc l'expérimentateur n'y a pas accès. Sinon la simulation n'aurait aucun intérêt !
Supposons par exemple un énorme sac contenant 4200 haricots blancs et 5800 haricots rouges : la population. Appelons π la proportion d'haricots blancs présents dans le sac ($\pi = 0,42$).
Nous allons prélever des échantillons de 50 haricots de ce sac afin d'essayer d'estimer la proportion d'haricots de chaque sorte puis tester l'hypothèse $\pi = 0,5$ *vs* $\pi \neq 0,5$. Nous allons comparer deux modes de prélèvement :

— tirage simultané de 50 haricots,
— tirage successif avec remise de 50 haricots (tirage aléatoire simple).

```
#simulation d'echantillonnage d'une proportion (grands echantillons)
pop <- 1:10000 #les 4200 haricots blancs et 5800 haricots rouges
taille_ech <- 50 #taille d'un echantillon
nSims <- 10000 #le nombre de simulations
prop <- numeric(nSims) #pour stocker les proportions simulees
bars <- 40 #le nombre de subdivisions de l'histogramme
for (i in 1:nSims)
```

```
  {ech <- sample(pop, taille_ech) #echantillon exhaustif
  blancs <- length(ech[ech<4201]) #compte le nb d'haricots blancs
  prop[i] <- blancs / taille_ech  #et le stocke en frequence dans prop
  }
#histogramme
hist(prop, breaks = bars, xlab="Proportions d'haricots blancs",
     ylab="Effectifs", axes=FALSE,
     main=paste("Distribution des proportions"),
     col="grey", xlim=c(0.1,0.75),  ylim=c(0, nSims/4))
axis(side=1, at=seq(0,1, 0.05), labels=seq(0,1,0.05))
axis(side=2, at=seq(0,nSims/4, nSims/8),
     labels=seq(0,nSims/4, nSims/8), las=2)
```



Remarque 1-2-10 : La simple vue de l'histogramme suffit à nous convaincre que $\pi \neq 0,5$, ce qui évite d'avoir recours à un test : la conclusion est évidente !
Sauf que dans la pratique, il est irréel d'effectuer 10000 tests pour vérifier une théorie. Ce qui rend cette simulation amusante certes, mais uniquement d'un point de vue théorique. Le commande **R** mean(prop) nous donne une estimation de π :
> mean(prop)
[1] 0.419894

Convaincu ? Le lecteur s'amusera à modifier la taille de l'échantillon, le mode d'échantillonnage (avec remise par exemple), etc. Nous allons entrer un peu plus dans la démarche de simulation dès les sections suivantes, ce qui nous amène également à quelques notions théoriques et à la pratique d'une méthode : celle du *maximum de vraissemblance* (en complément).

1.3 Les $p-$valeurs

Ceci pourrait commencer par une question : "Où sont les extrêmes ?" : à droite ? à gauche ? Ou des deux côtés en valeur absolue ? Vous n'avez pas tout compris ? Alors voici ce quiz :

Quiz 4 : Une usine produit des sodas dont le pH moyen est estimé à 2,7. On suppose la répartition des pH normale et d'écart-type $\sigma = 0,25$. Lors d'un contrôle qualité, on prélève un E.A.S de 20 sodas. Soit $\bar{X}$ la variable aléatoire qui à tout E.A.S de 20 sodas associe leur pH moyen. Notre échantillon nous donne un pH moyen de $\bar{x} = 2,57$. Nous posons pour hypothèse nulle H_0 : "le PH est conforme au pH indiqué" et pour hypothèse alternative H_1 : "le pH est strictement inférieur au pH indiqué".

1. (a) C'est un test bilatéral (b) C'est un test unilatéral à droite (c) C'est un test unilatéral à gauche (d) C'est un test à la con !

2. Sous H_0, la loi de $\bar{X}$ est une loi normale de paramètres $(\mu_0; \sigma_{0,n})$ égaux à :
   (a) (2,7 ; 0,25) (b) (2,57 ; 0,25) (c) (2,57 ; 0,056) (d) (2,7 ; 0,056)
3. Sous H_0, $P(\bar{X} \leq 2,57)$ est égale à :
   (a) 0,0101 (b) 0,3015 (c) Autre (d) 0,5
4. Sous H_0, on prend pour variable de décision $U = \dfrac{\bar{X} - \mu_0}{\sigma_{0,n}} \hookrightarrow \mathcal{N}(0;1)$. u_{obs} est égal à :
   (a) $-2,3255$ (b) $-1,8556$ (c) 2,3255 (d) 1,3642
5. $P(\bar{X} \leq 2,57)$ et $P(U \leq u_{obs})$ sont égaux.
   (a) Vrai (b) Faux
6. $P(U \leq u_{obs}) < \alpha$ revient à rejeter H_0.
   (a) Faux (b) Vrai
7. On prend pour hypothèse alternative H_1 : "le pH est différent du pH indiqué". Au risque de première espèce α de 5% et en prenant $U = \dfrac{\bar{X} - \mu_0}{\sigma_{0,n}} \hookrightarrow \mathcal{N}(0;1)$ comme variable de décision sous H_0, la zone de non rejet de H_0 est :
   (a) $[-2,58; 2,58]$ (b) $[-1,96; 1,96]$ (c) $[-1,645; 1,645]$ (d) Autre
8. Quelles données reviennent à rejeter H_0 ? (plusieurs réponses possibles)
   (a) $u_{obs} \notin [-1,96; 1,96]$ (b) $P(|U| > |u_{obs}|) < \alpha$ (c) $P(|U| > |u_{obs}|) \geq \alpha$

Nous supposons que nous cherchons à vérifier la **conformité d'une moyenne** μ à une norme théorique μ_0.
H_0 : "il n'y a pas de différence significative entre la moyenne observée et la moyenne théorique", ce qui se traduit formellement par $\mu = \mu_0$. Nous distinguerons deux cas :

1. L'écart-type σ du paramètre observé dans la population est connu,
2. L'écart-type σ du paramètre observé dans la population est inconnu.

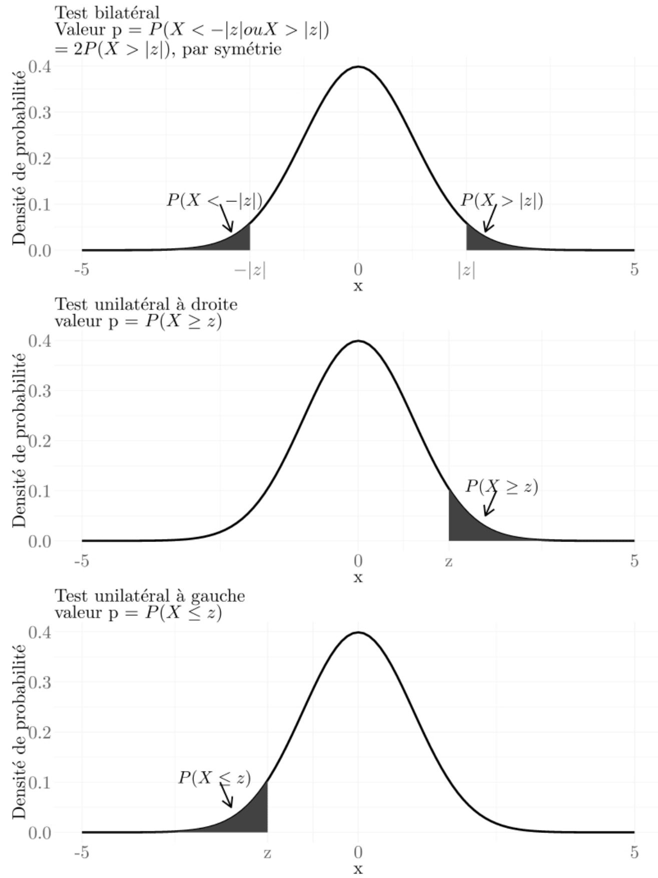
Ces deux situations amènent à utiliser deux types de variables de décision différentes que nous résumons dans le tableau ci-dessous.

σ connu	H_1	Variable de décision sous H_0	$p-$valeurs
	$\mu \neq \mu_0$		$P(\lvert U\rvert > \lvert u_{obs}\rvert)$
$Z-$test	$\mu > \mu_0$	$Z = U = \dfrac{\bar{X} - \mu}{\sigma/\sqrt{n}}$	$P(U > u_{obs})$
	$\mu < \mu_0$		$P(U < u_{obs})$

σ inconnu	H_1	Variable de décision sous H_0	$p-$valeurs
	$\mu \neq \mu_0$		$P(\lvert T\rvert > \lvert t_{obs}\rvert)$
$T-$test	$\mu > \mu_0$	$T = \dfrac{\bar{X} - \mu}{\hat{S}/\sqrt{n}}$	$P(T > t_{obs})$
	$\mu < \mu_0$		$P(T < t_{obs})$

La variable T suit une loi de Student[3] à $n-1$ degrés de liberté (ddl). T et $Z = U$ (la loi normale centrée-réduite) ont des fonctions de densités symétriques par rapport à l'axe des ordonnées. Ainsi, si nous notons X pour U ou T indifféremment :

3. détaillé dans la partie compléments



images issues de Google

Nous pouvons résumer la notion de $p-$valeur sous cette forme imagée : $\boxed{P_{H_0}(D+)}$, où $D+$ désigne l'événement "X prend D ou une valeur plus extrême" comme résumé graphiquement ci-dessus.
Autrement dit, nous évaluons sous l'hypothèse nulle H_0 la vraisemblance de l'événement $D+$. Alors qu'en fait, nous cherchons, au vu de nos données, à évaluer $\underline{P_{D+}(H_0)}$. Nous discuterons de ceci dans un prochain article sur les statistiques bayésiennes.

Retenons le fait suivant : **si la $p-$valeur $< \alpha$, alors nous rejetons** H_0.

Ce qui nous amène à la rédaction d'un test en quatre étapes :

1. Formulation des hypothèses H_0 et H_1
2. On se fixe le seuil de risque de première espèce α
3. Sous H_0, on donne la loi de probabilité d'une variable de décision X
4. En utilisant les données de notre échantillon et X, on calcule la $p-$valeur. Si cette dernière est strictement inférieure à α, on rejette H_0 et on conclut que les données observées ne sont probablement pas que "du bruit". Il faut chercher une explication ailleurs que dans une simple fluctuation d'échantillonnage.

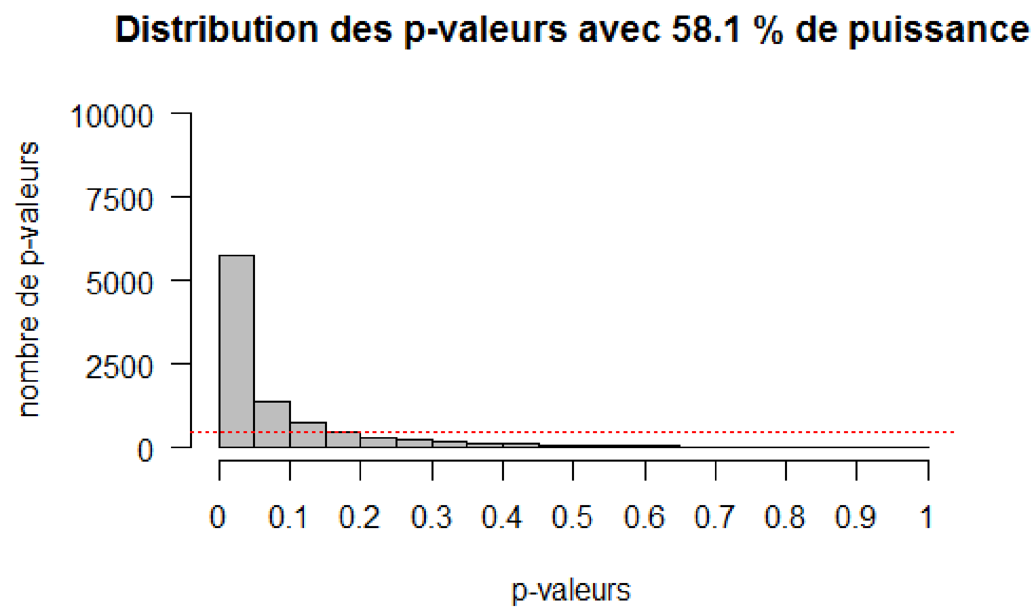
Pour le moment, nous allons faire une constatation fondamentale : **la distribution des $p-$valeurs est directement reliée à la puissance du test**.
Une petite simulation va nous aider à y voir plus clair [4] ...

4. le script qui suit est inspiré de celui de Daniel Lakens de l'université technique d'Eindhoven (NL)

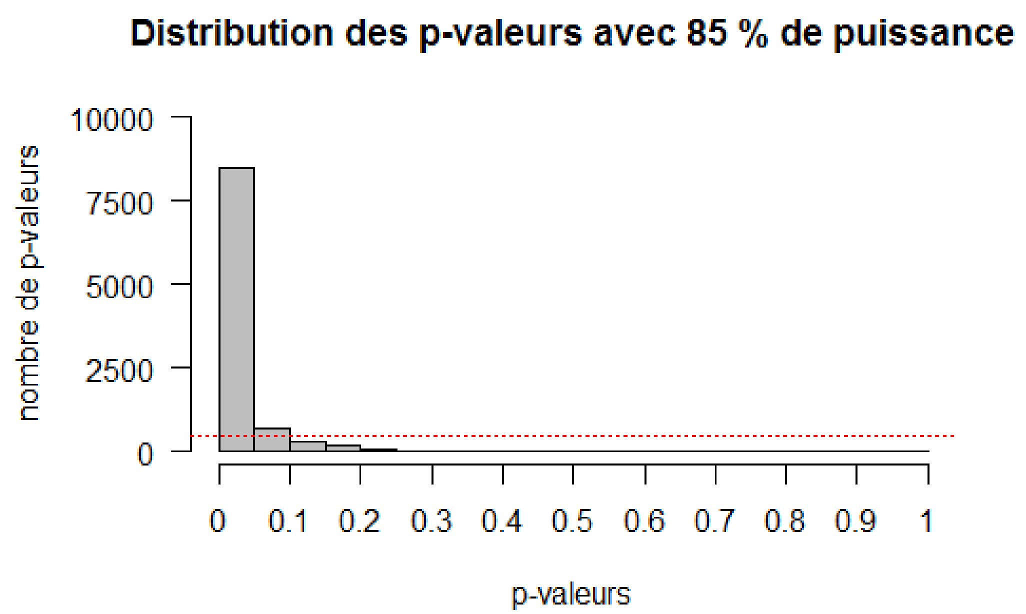
Considérons une usine de sodas. Le pH moyen des bouteilles de cette production est supposé égal à 2,5. La réalité (inconnue du contrôleur qualité) est que le pH des sodas de cette production est distribué selon une loi normale de moyenne $\mu = 2,3$ et d'écart-type $\sigma = 0,32$.

```
#Utilisation du package pwr pour calculer la puissance du test
if(!require(pwr)){install.packages('pwr')}
library(pwr)
#Desactive la notation scientifique (1.05e10)
options(scipen=999)
nSims <- 10000 #nombre de simulations
M <- 2.3   #moyenne de la population dont est issu l'echantillon
taille_ech <- 14    #taille de l'echantillon
SD <- 0.32 #ecart-type des donnees simulees
p <- numeric(nSims) #vecteur de stockage des p-valeurs
bars <- 20
for(i in 1:nSims){ #pour chaque experience simulee
  x <- rnorm(n = taille_ech, mean = M, sd = SD) #donnees crees
  z <- t.test(x, mu = 2.5) #applique un t-test de conformite a mu
  p[i] <- z$p.value #calcule la p-valeur et la stocke dans p
}
#Calcul pratique de la puissance
print("puissance du test : ")
print(sum(p < 0.05)/nSims) #puissance empirique
#Calcul de la puissance avec la package pwr
power <- pwr.t.test(d=(M-2.5)/SD, n=taille_ech, sig.level=0.05,type="one.sample",
                    alternative="two.sided")$power
#Trace de la distribution pratique des p-valeurs
op <- par(mar = c(5,7,4,4))
hist(p, breaks=bars, xlab="p-valeurs", ylab="nombre de p-valeurs\n", axes=FALSE,
     main=paste("Distribution des p-valeurs avec",round(power*100, digits=1),
                "% de puissance"),
     col="grey", xlim=c(0,1),  ylim=c(0, nSims))
axis(side=1, at=seq(0,1, 0.1),
     labels=seq(0,1,0.1))
axis(side=2, at=seq(0,nSims, nSims/4), labels=seq(0,nSims, nSims/4), las=2)
```

Avec une taille d'échantillon de 14, on obtient une puissance de l'ordre de 58% et une distribution empirique des $p-$valeurs qui a pour allure :



Avec une taille d'échantillon de 25 (prendre taille_ech < − 25), on obtient une puissance de l'ordre de 85% et une distribution empirique des $p-$valeurs qui a pour allure :



Remarque 1-3-1 : La taille de l'échantillon augmente donc sans conteste la puissance du test. Mais allons plus loin dans la compréhension . . .

Quiz 5 : Le but de ce quiz est de tester empiriquement les paramètres qui jouent sur la puissance d'un test et d'étudier les deux cas :

— il y a vraiment un effet significatif (H_0 est fausse)

— il n'y a pas d'effet significatif (H_0 est vraie)

1. Obtenir une $p-$valeur strictement inférieure à α à l'issue d'un test signifie que :
   (a) H_1 est vraie (b) H_0 est fausse (c) Sous H_0, les données obtenues sont improbables (d) Sous H_1, les données obtenues sont vraisemblables
2. Considérons le t-test de conformité à une moyenne de l'exemple précédent. Pour augmenter sa puissance, on peut :
   (a) diminuer $|M - 2,5|$ (b) Augmenter $|M - 2,5|$ (c) augmenter SD (d) diminuer SD
3. Réaffecter à M la valeur 2,5 dans le script précédent (ce qui revient à supposer H_0 vraie). La distribution des $p-$valeurs semble :
   (a) être uniforme (b) être normale (c) être exponentielle (d) autre
4. La puissance obtenue à la question 3 est indépendante de la taille de l'échantillon.
   (a) Vrai (b) Faux
5. Le défaut de puissance d'un test conduit plus souvent à ne pas rejeter H_0.
   (a) Vrai (b) Faux

Le prochain quiz devrait vous permettre de ne pas tomber dans un jeu d'interprétations erronées. Surtout, pensez à la définition d'une $p-$valeur !

Quiz 6 : Vrai ou faux ? Mais comprenez pourquoi !

1. Lors d'un test statistique, on obtient une $p-$valeur de 0,03. Ceci signifie que la probabilité que H_0 soit vraie est de 0,03.
   (a) Vrai (b) Faux
2. Obtenir une grande $p-$valeur est une preuve en faveur de H_0.
   (a) Vrai (b) Faux
3. Rejeter l'hypothèse nulle car on a obtenu une $p-$valeur $p \leq 0,05$ signifie que la probabilité de se tromper est de 5%.
   (a) Vrai (b) Faux
4. Quand la même hypothèse nulle est testée dans différentes études et qu'aucune ou une minorité des tests sont statistiquement signifiants (tout ou presque tous les p $>$ 0,05), ceci est une preuve globale en faveur de H_0.
   (a) Vrai (b) Faux
5. La notion de $p-$valeur est une notion fréquentiste.
   (a) Vrai (b) Faux

1.4 Survol de quelques tests classiques en BTSA

Avant de poursuivre plus avant, effectuons une synthèse partielle de ce que nous avons appris précédemment : un test d'hypothèse consiste à confronter deux hypothèses incompatibles H_0 et H_1 portant sur un paramètre de la population (cas des tests paramétriques étudiés ici). En supposant H_0 vraie, nous regardons, avec un seuil d'erreur fixé à l'avance (celui de rejeter H_0 à tort), si nos données nous amènent à rejeter H_0 ou à la conserver, sans que ceci ne prouve rien sur sa véracité. Un pseudo-raisonnement par l'absurde !

1.4.1 Tests de conformité

Nous allons présenter dans ce paragraphe trois situations classiques :

1. le test de conformité d'une moyenne à une norme (déjà abordé auparavant),
2. le test de conformité d'une variance,
3. et le test de conformité d'une proportion (cas des grands échantillons).

Une hypothèse (éventuellement à vérifier) est faite pour les deux premiers tests : le caractère quantitatif étudié est supposé réparti normalement.

1.4.1.1 : Test de conformité d'une moyenne ($H_0 : \mu = \mu_0$).

σ connu	H_1	Variable de décision sous H_0
$Z-$test	$\mu \neq \mu_0$ $\mu > \mu_0$ $\mu < \mu_0$	$Z = \dfrac{\bar{X} - \mu}{\sigma/\sqrt{n}}$
σ inconnu	**H_1**	**Variable de décision sous H_0**
$T-$test	$\mu \neq \mu_0$ $\mu > \mu_0$ $\mu < \mu_0$	$T = \dfrac{\bar{X} - \mu}{\hat{S}/\sqrt{n}}$

$\hat{S}$ est la variable aléatoire qui à tout EAS de taille n lui associe son écart-type corrigé $\hat{s}$ (cf paragraphe suivant).

Exemple 6 : Un scientifique souhaite comparer le QI moyen des habitants d'une ville bretonne avec le QI moyen de la population française, estimé à 100. Pour ceci, il prélève un échantillon (assimilé à un échantillon aléatoire simple) de 20 individus de cette ville. Il constate que le QI moyen observé est de 106 et l'*écart-type corrigé*[5] de 15,2. Peut-on conclure à une différence significative entre le QI des habitants de cette ville bretonne et celui de la population française ?

Nous allons à travers cet exemple détailler deux manières de traiter un test statistique classique (avec les zones de rejet et les p-valeurs). La situation choisie est celle où l'écart-type de la population est inconnue (réaliste). Le cas d'école est traité en exercice.

5. $\hat{s} = \sqrt{\dfrac{1}{n-1}\displaystyle\sum_{i=1}^{n}(x_i - \bar{x})^2}$, noté *sd* (standard deviation) chez les anglo-saxons

Méthode 1 : Nous allons raisonner en 6 étapes.

Soit μ le QI moyen d'un habitant de cette ville bretonne.

Étape 1 : formulation des hypothèses.
$\begin{cases} H_0 : \mu = 100 \\ H_1 : \mu \neq 100 \text{ (test bilatéral)} \end{cases}$

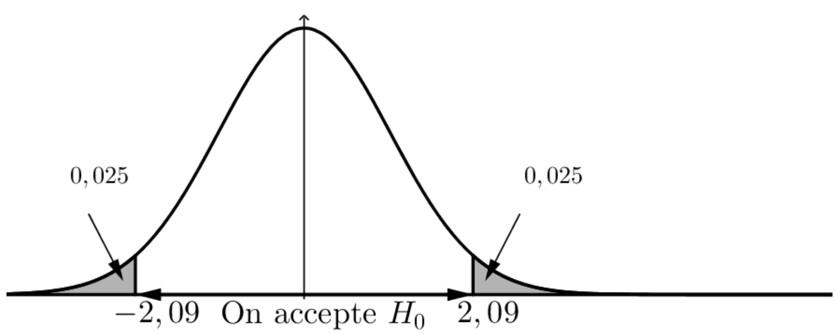
Étape 2 : on se fixe le risque de première espèce α.
On choisit ici $\alpha = 5\%$.

Étape 3 : on détermine la variable de décison en supposant H_0 vraie.
Il est implicite ici, mais ceci mérite d'être précisé, que le QI des français est distribué normalement dans la population. L'écart-type σ de la population étant inconnu et l'échantillon de taille $20 < 30$ nous incite à utiliser la loi de Student.
Soit $\bar{X}$ (resp. $\hat{S}$) la variable aléatoire qui à tout EAS de 20 individus de cette ville bretonne associe le QI moyen (resp. l'écart-type corrigé). Alors :

$$T = \frac{\bar{X} - 100}{\frac{\hat{S}}{\sqrt{20}}} \hookrightarrow \text{ Student à 20-1} = 19 \text{ ddl}$$

Remarque : La variable de décision T s'exprime aussi sous la forme $T = \dfrac{\bar{X} - \mu}{S/\sqrt{n-1}}$, où S^2 est la variable aléatoire qui à tout E.A.S de taille n lui associe sa variance (non corrigée).

Étape 4 : on détermine les zones de rejet (car test bilatéral) et de non-rejet de H_0.
La détermination des seuils critiques se fait à l'aide d'un software ou d'une table de Student.



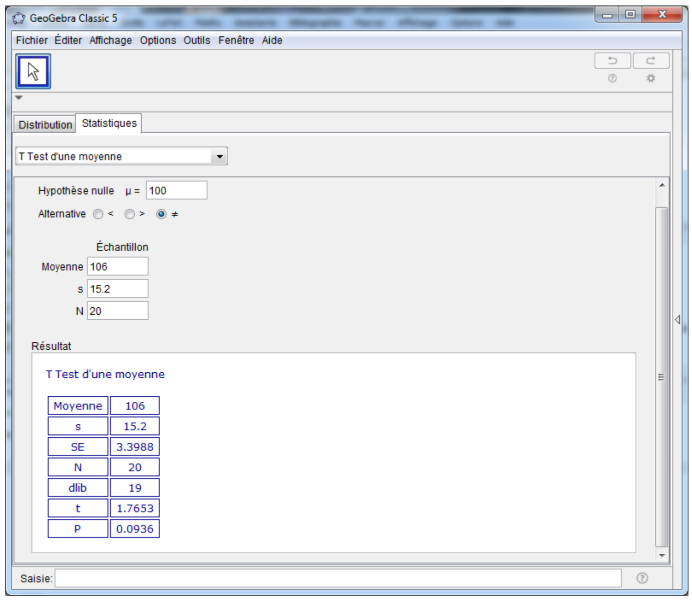
Étape 5 : on énonce la règle de décision.

— Si $t_{\text{obs}} \in [-2,09; 2,09]$, on ne rejette pas H_0,
— Si $t_{\text{obs}} \notin [-2,09; 2,09]$, on rejette H_0.

Étape 6 : on calcule la statistique observée et on conclut à l'aide de la règle de décision.
$t_{\text{obs}} = \dfrac{106 - 100}{\frac{15,2}{\sqrt{20}}} = 1,765 \in [-2,09; 2,09]$.

Donc au risque de première espèce de 5%, on ne rejette pas H_0. On impute la différence observée aux fluctuations d'échantillonnage et on peut affirmer que le QI moyen des habitants de cette ville bretonne n'est pas significativement différent du QI moyen en France.

<u>**Méthode 2**</u> : Elle se base sur l'utilisation des $p-$valeurs, données par tous les softwares en statistiques.
Les étapes 1, 2 et 3 sont identiques.
L'étape 4 consiste à calculer la $p-$valeur à l'aide des données de notre échantillon et de la nature du test.
Ici, il s'agit donc de calculer $P(|T| \geq t_{\text{obs}})$ *i.e* $P(|T| \geq 1,765)$, où $T \hookrightarrow$ Student à 19 ddl.
L'utilisation de Geogebra nous amène à $p = 0,0936$. Cette valeur dépasse 0,05 donc nous ne rejettons pas l'hypothèse nulle au risque de première espèce de 5%.



Nous conservons donc H_0 avec un risque β inconnu.

Exercice 2 : Un boulanger produit des baguettes dont la masse affichée est 250g. Un client suspicieux relève pendant deux semaines les masses des baguettes (exprimée en grammes) qu'il achète : 249, 251, 248, 247, 251, 250, 251, 248, 249, 252, 250, 251, 249, 249. Nous supposons de plus que les masses sont réparties normalement dans la production.

1. Calculer la moyenne, l'écart-type et l'écart-type corrigé de cet échantillon.
2. En supposant l'écart-type de la production égal à 2g, peut-on supposer, au risque de première espèce de 5%, que la masse moyenne μ d'une baguette est inférieure aux 250g affichés.
3. Même question, mais en supposant l'écart-type de la production inconnu.

1.4.1.2 : Test de conformité d'une variance $(H_0 : \sigma^2 = \sigma_0^2)$

Soient $x_1, x_2, \ldots, x_n$ n observations d'une variable aléatoire quantitative. Nous définissons :

— La moyenne des x_i par $\bar{x} = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} x_i$,

— la somme des carrés des écarts à la moyenne par $\text{SCE} = \displaystyle\sum_{i=1}^{n}(x_i - \bar{x})^2$

La variance s^2 de notre échantillon est égale à $\frac{\text{SCE}}{n} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$. C'est une réalisation de la variable aléatoire S^2 qui à tout E.A.S de taille n associe sa variance.
Problème : la variable aléatoire S^2 est biaisée ! En effet, nous pouvons prouver que $E(S^2) = \frac{n-1}{n}\sigma^2$, où σ^2 est la variance du caractère dans la population. Par linéarité de l'espérance, nous définissons $\hat{S}^2 = \frac{n}{n-1}S^2$, de sorte que $E(\hat{S}^2) = \sigma^2$: $\hat{S}^2$ **est ainsi un estimateur non biaisé de la variance** σ^2 et $\hat{s}^2$ une de ses réalisations. Nous avons $\boxed{nS^2 = (n-1)\hat{S}^2}$.

Remarquons que pour "passer de $Z = U$ à T" dans la variable de décision du test de conformité d'une moyenne, nous avons remplacé σ (connu dans le premier cas) par son estimateur $\hat{S}$. Attention, $\hat{S}$ n'est pas un estimateur sans biais de l'écart-type σ.

ATTENTION AUX NOTATIONS : Dans de nombreux ouvrages ou logiciels comme Geogebra, l'écart-type corrigé de l'échantillon (parfois appelé écart-type expérimental) est noté s et non $\hat{s}$. C'est *sd* (standard deviation) sous **R** ou Python.
Aussi, au moment de saisir les données dans ces logiciels, c'est donc l'écart-type corrigé qu'il faudra indiquer !

Théorème 1.4.1.2 (a) : $\frac{(n-1)\hat{S}^2}{\sigma^2} = \frac{nS^2}{\sigma^2} \hookrightarrow \chi^2$ à $n-1$ degrés de liberté.

Ce résultat va nous permettre de construire un intervalle de confiance de la variance pour un risque α fixé. Mais comprenez-vous pourquoi $\frac{nS^2}{\sigma^2}$ suit la loi du χ^2 à $n-1$ degrés de liberté ?

Théorème 1.4.1.2 (b) : Un intervalle de confiance au risque de confiance $1-\alpha$ de la variance σ^2 d'un caractère de la population est $\left[\frac{ns^2}{\chi^2_{n-1;1-\alpha/2}}; \frac{ns^2}{\chi^2_{n-1;\alpha/2}}\right] = \left[\frac{(n-1)\hat{s}^2}{\chi^2_{n-1;1-\alpha/2}}; \frac{(n-1)\hat{s}^2}{\chi^2_{n-1;\alpha/2}}\right]$.

Quiz 7 : Intervalle de confiance et test bilatéral.
Un technicien agricole souhaite réaliser une étude sur les masses des porcelets nés vivants dans une ferme expérimentale. Un échantillon aléatoire simple indépendant de 16 porcelets nés le mois précédent est prélevé dans la ferme. On suppose que les masses des porcelets nés vivants sont distribuées normalement.
Les masses mesurées, en kilogramme, sont les suivantes :
1,85 1,36 1,75 0,88 1,72 1,99 1,40 1,40 1,49 1,22 1,43 1,65 1,38 1,65 1,17 1,48

1. La variance de cet échantillon est égale à 10^{-4} près à :
   (a) 0,0762 (b) 0,0811 (c) 0,0765 (d) 0,0717
2. La borne inférieure de l'intervalle de confiance de σ^2 au seuil de confiance de 95% est égale à 10^{-3} près à :
   (a) 0,052 (b) 0,041 (c) 0,082 (d) 0,182
3. On teste l'hypothèse $H_0 : \sigma^2 = 0,204$ contre $H_1 : \sigma^2 \neq 0,204$. Rejette-t-on H_0 au risque α de 5% ?
   (a) Oui (b) Non

Exemple 7 : En vue d'estimer le rendement d'une culture de blé tendre, un technicien agricole a pesé, en kg, les grains récoltés sur 16 parcelles de $2m^2$ choisies au hasard de manière indépendante.

Les valeurs relevées sont les suivantes :
1,32 1,07 1,26 1,02 1,17 1,05 1,27 1,12 1,21 1,15 1,27 1,20 1,23 1,10 1,06 1,14
Peut-on affirmer, au vu de cet échantillon, que l'écart-type du rendement est significativement supérieur à 3,5 q/ha ?

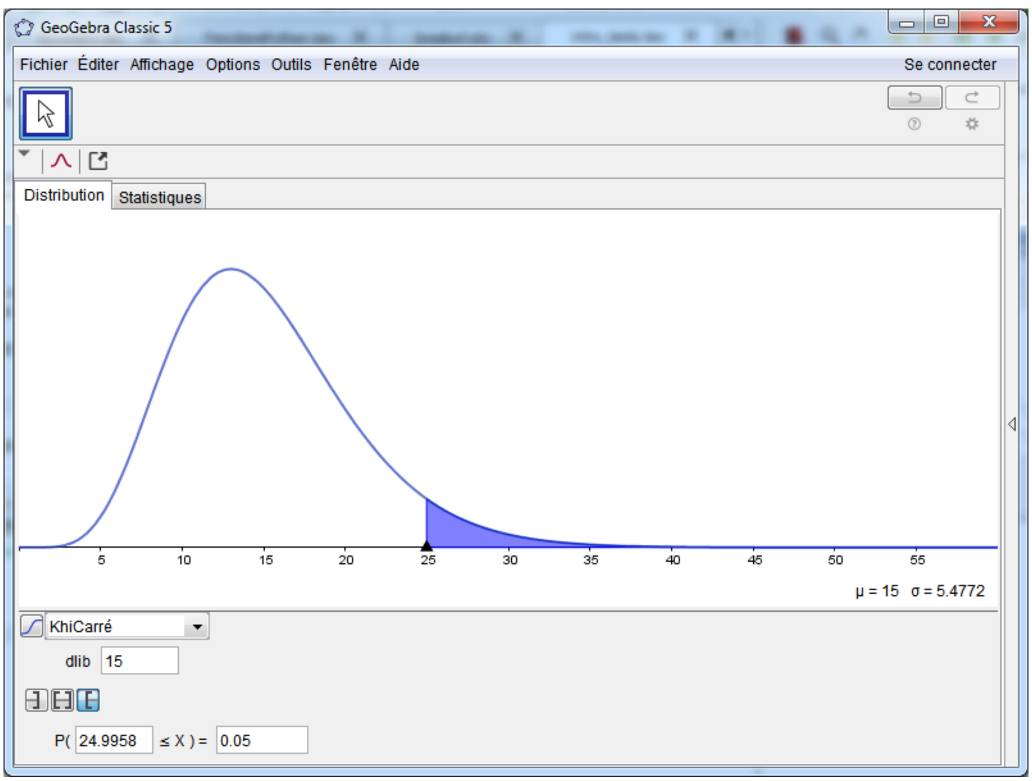
Solution : L'énoncé nous incite à poser $H_0 : \sigma \leq 3,5$ *vs* $H_1 : \sigma > 3,5$, qui sont deux hypothèses composites. Comme la loi de la statistique de décision doit être parfaitement déterminée sous H_0, nous la reformulons comme $H_0 : \sigma = 3,5$ (à comprendre comme la borne supérieure de l'écart-type du rendement n'est pas significativement différente de 3,5 q/ha).
Reformulons encore notre hypothèse nulle, puisque comparer des écarts-type est équivalent à comparer des variances (quantités toutes positives) :

Étape 1 : formulation des hypothèses $\begin{cases} H_0 : \sigma^2 = 3,5^2 = 12,25 \\ H_1 : \sigma^2 > 12,25 \text{ (test unilatéral à droite)} \end{cases}$

Étape 2 : on se fixe le risque de première espèce α à 5 %.

Étape 3 : on détermine la variable de décison en supposant H_0 vraie.
Soit S^2 la variable aléatoire qui à tout EAS de 16 parcelles associe la variance du rendement.
$K = \dfrac{16S^2}{12,25} \hookrightarrow \chi^2$ à 15 ddl.

Étape 4 : on détermine les zones de rejet et de non-rejet de H_0. On se sert pour trouver le seuil critique d'une table statistique du khi-deux ou d'un software.



Étape 5 : on énonce la règle de décision.
- Si $k_{\text{obs}} \leq 25$, on ne rejette pas H_0,
- Si $k_{\text{obs}} > 25$, on rejette H_0.

Étape 6 : on calcule la statistique observée et on conclut à l'aide de la règle de décision.
Et là, oh là, oh là, oh là là ! Pour une fois c'est ici que ça devient délicat ! Toutes les données

que nous observons sur nos parcelles doivent être exprimées en quintaux par hectare ! Nos observations sont $s = 0{,}088$ kg/2m$^{\mathbf{2}}$, soit après conversion $s = 4{,}4$ q/ha.
D'où $k_{\text{obs}} = \dfrac{16 \times 4,4^2}{12,25} = 25,29 > 25$.
Donc au risque de première espèce de 5%, on rejette H_0. On impute la différence observée à une cause autre que les fluctuations d'échantillonnage et on peut affirmer que l'écart-type du rendement est significativement supérieur à 3,5 q/ha.

Question : si l'on avait utilisé un test bilatéral, la conclusion aurait-elle été identique ? Proposer une explication.

Exercice 3 : Dans une exploitation, on a relevé les masses d'un lot de 24 agneaux lors d'une pesée. On suppose que les masses sont distribuées normalement.
Les relevés, en kg, sont les suivants :
21,3 18,6 20,5 19,6 20,2 20,3 19,6 20,8
19,0 23,1 16,8 21,0 20,1 16,8 21,4 18,3
19,7 22,1 20,9 17,3 21,6 16,4 20,3 20,1

1. Donner une estimation ponctuelle non biaisée de la variance des masses des agneaux,
2. Donner une estimation par intervalle de confiance de la variance des masses des agneaux au risque de 5%,
3. Peut-on affirmer, au vu de cet échantillon et au risque de 5%, que l'écart-type des masses des agneaux est significativement différent de 2,5 kg ?

1.4.1.3 : Test de conformité d'une proportion $(H_0 : \pi = \pi_0)$

Conformément au programme, nous traiterons le cas des grands échantillons (n>30). Le cas des petits échantillons peut être résolu avec la loi exacte que nous avons déjà rencontrée au quiz 3. Par souci de détail, nous comparerons loi exacte et loi approchée de la variable de décision de notre test.
Considérons un caractère que présentent certains individus dans une population : couleur des yeux, groupe sanguin, taille, etc. On note π la proportion d'individus présentant ce caractère. Dans certains cas, on cherche à savoir si cette proportion π est conforme à une norme π_0.

Théorème 1.4.1.3 (a) : Soit F la variable aléatoire qui à tout échantillon aléatoire simple de taille n associe la proportion p (parfois notée f) d'individus présentant le caractère observé.

1. $F = \dfrac{X}{n}$, où $X \hookrightarrow \mathcal{B}(n, \pi)$,
2. $E(F) = \pi$,
3. $Var(F) = \dfrac{\pi(1-\pi)}{n}$.

Théorème 1.4.1.3 (b) : Si $n > 30$, $n\pi \geq 5$ et $n(1-\pi) \geq 5$, alors on peut "approcher" la loi de F par la loi normale de moyenne $\mu = \pi$ et d'écart-type $\sigma = \sqrt{\dfrac{\pi(1-\pi)}{n}}$.
Ainsi, $U = \dfrac{F - \pi}{\sqrt{\dfrac{\pi(1-\pi)}{n}}} \hookrightarrow \mathcal{N}(0; 1)$.

Théorème 1.4.1.3 (c) : Si $n > 30$, $np \geq 5$ et $n(1-p) \geq 5$, alors :

1. Une estimation ponctuelle de π est p (valable sans les conditions précitées),
2. Un intervalle de confiance de π au seuil de confiance $1-\alpha$ est

$$\left[p - u_{1-\alpha/2}\sqrt{\frac{p(1-p)}{n}}; p + u_{1-\alpha/2}\sqrt{\frac{p(1-p)}{n}}\right]$$

Quiz 8 : Pour chaque item, une seule réponse est possible.
Une affection atteint 11% des ovins en France. Un vétérinaire teste un nouveau vaccin sur un troupeau de 150 ovins. Parmi les bêtes vaccinées, on dénombre 14 bêtes atteintes par cette affection. Le but est de savoir si le vaccin est efficace *i.e* contribue à baisser le taux de contamination.

1. Une estimation ponctuelle p de la proportion de bêtes vaccinées atteintes par l'affection est :
   (a) 0,0867 (b) 0,0933 (c) 0,1111 (d) 0,1215
2. La borne supérieure de l'intervalle de confiance, au seuil de risque de 10%, de la proportion de bêtes vaccinées atteintes par l'affection est :
   (a) 0,1324 (b) 0,0542 (c) 0,1399 (d) 0,1546
3. Au vu de la question, l'hypothèse alternative H_1 est une hypothèse :
   (a) bilatérale (b) unilatérale à droite (c) unilatérale à gauche (d) bancale
4. On prend X comme variable de décision (compteur de bêtes infectées sous H_0) et un risque α de 5%. On rejette H_0 dans le cas où $x = 150p$ appartient à :
   (a) $[\![0; 9]\!]$ (b) $[\![0; 10]\!]$ (c) $[\![10; 141]\!]$ (d) $[\![141; 150]\!]$
5. Avec X comme variable de décision, on rejette H_0.
   (a) Vrai (b) Faux
6. On prend U comme variable de décision. La valeur u_{obs} prise par U pour notre échantillon est :
   (a) $-0,7435$ (b) 0,6537 (c) 1,3469 (d) $-0,6537$ (e) 0,7435
7. Avec U comme variable de décision, on rejette H_0.
   (a) Vrai (b) Faux

Exemple 8 : Le taux de réussite moyen au concours d'emplumeurs d'oreillers est de 23%. On prélève dans la grande région du sud-ouest de la Canardie un EAS de 500 étudiants à ce concours. 122 d'entre eux sont reçus. Peut-on, au vu de cet échantillon, et au risque de première espèce de 5%, affirmer que le taux de réussite à ce concours des étudiants de cette région est différent de la moyenne nationale ?

Solution : soit π la proportion de reçus au concours des futurs d'emplumeurs d'oreillers en Canardie. On teste $H_0 : \pi = 0,23$ *vs* $H_1 : \pi \neq 0,23$.
On rejette H_0 si $|u_{obs}| = \dfrac{|f - 0,23|}{\sqrt{0,23 \times 0,77/500}} > 1,96$. Ici, $|u_{obs}| = 0,744$, donc au vu de cet échantillon et au risque de première espèce de 5%, on ne rejette pas H_0 : on peut affirmer que le taux de réussite à ce concours des étudiants de la région n'est pas significativement différent du taux national.
Remarque : on aurait pu raisonner à l'aide d'un intervalle de confiance (exercice).

1.4.2 Tests de comparaison

Nous allons présenter dans ce paragraphe trois situations classiques :

1. le test de comparaison de deux proportions (cas des grands échantillons).
2. le test de comparaison de deux variances,
3. le test de comparaison de deux moyennes.

Une hypothèse (éventuellement à vérifier) est faite pour les deux derniers tests : le caractère quantitatif étudié est supposé réparti normalement.

1.4.2.1 : Test de comparaison de deux proportions $(H_0 : \pi_1 = \pi_2)$.

Nous travaillerons dans le cas de "grands échantillons". Mais raisonnons d'abord en terme de lois exactes.
Considérons deux populations desquelles nous prélevons deux échantillons aléatoires simples *indépendants* de tailles respectives n_1 et n_2. Appelons X_1 (resp. X_2) la variable aléatoire qui à tout EAS de n_1 (resp. n_2) individus de la population 1 (resp. population 2), lui associe le nombre d'individus présentant le caractère observé.
X_1 (resp. X_2) suit la loi binomiale de paramètres n_1 et π_1 (resp. n_2 et π_2).

Mais comme nous cherchons à comparer deux proportions, il nous faut un estimateur de $\pi_1 - \pi_2$.
Posons donc $F_1 = \dfrac{X_1}{n_1}$ et $F_2 = \dfrac{X_2}{n_2}$ qui sont deux estimateurs (sans biais) des proportions inconnues π_1 et π_2.
Sous H_0, nous avons par linéarité de l'espérance : $E(F_1) - E(F_2) = E(F_1 - F_2) = \pi_1 - \pi_2 = 0$.
Par indépendance de F_1 et de F_2, nous avons : $Var(F_1 - F_2) = Var(F_1) + Var(F_2) = \dfrac{Var(X_1)}{n_1^2} + \dfrac{Var(X_2)}{n_2^2} = \dfrac{\pi_1(1-\pi_1)}{n_1} + \dfrac{\pi_2(1-\pi_2)}{n_2} = \pi(1-\pi)\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)$, où π est la valeur commune mais inconnue de π_1 et de π_2.

Ainsi, si l'on pose $F = F_1 - F_2$, alors sous H_0 : $E(F) = 0$ et $\sigma(F) = \sqrt{\pi(1-\pi)\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)}$.
Hélas, π est encore inconnue ! Il va falloir l'estimer aussi.
L'idée est de se servir d'une "proportion pondérée" tout comme on calcule une moyenne pondérée : on estime π par $\hat{\pi} = \dfrac{n_1\pi_1 + n_2\pi_2}{n_1 + n_2}$.

Sous H_0, F est centrée. Réduisant F en remplaçant π par $\hat{\pi}$, on a comme variable de décision

$$D = \frac{F_1 - F_2}{\sqrt{\hat{\pi}(1-\hat{\pi})\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)}}$$

Sous les conditions standard $n_1, n_2 \geq 30$, $n_1\hat{\pi} \geq 5$, $n_1(1-\hat{\pi}) \geq 5$, $n_2\hat{\pi} \geq 5$, $n_2(1-\hat{\pi}) \geq 5$, on peut approximer D par la loi normale centrée-réduite Z.

Sous H_0 : $\pi_1 = \pi_2$ et les conditions standard énoncées, la variable de décision sera :

$$\boxed{Z = \frac{F_1 - F_2}{\sqrt{\hat{\pi}(1-\hat{\pi})\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)}} \hookrightarrow \mathcal{N}(0,1)}$$

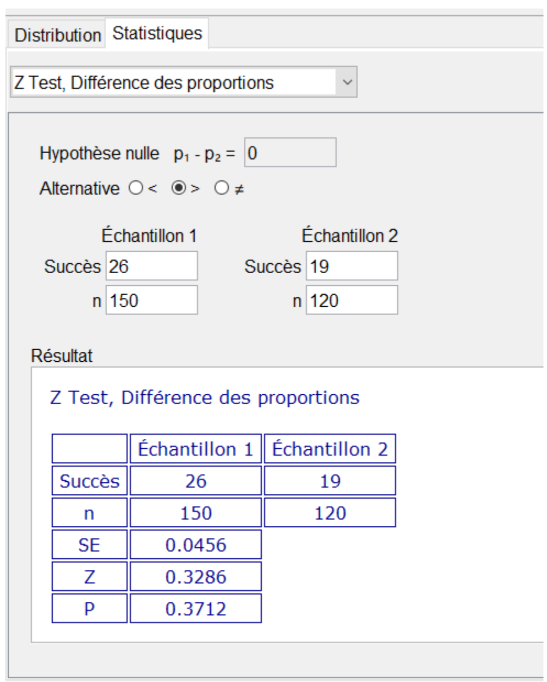
Exemple 9 : Deux populations homogènes de lapins ont été exposées à la redoutable bactérie *Bacillus cretinisus*. Mais le professeur Longuoreille et son équipe ont conçu deux vaccins censés protéger nos Leporidae (famille des lapins) préférés.

— Avec le vaccin 1, 26 lapins sur 150 de la population 1, ont été infectés.
— Avec le vaccin 2, 19 lapins sur 120 de la population 2, ont été infectés.

On suppose les échantillons de lapins vaccinés indépendants.
Au vu de cet échantillon, et au risque α de 5%, peut-on affirmer que le second vaccin est plus efficace que le premier ?

Solution : Appelons π_1 (resp. π_2) la proportion de lapins de la population 1 (resp. de la population 2) vaccinés mais qui vont tomber malades.
Remarquons que nous ne connaissons ni π_1, ni π_2 car seuls une petite partie de chaque population (nos deux échantillons) ont reçu ce vaccin.
Au vu de notre énoncé, nous allons tester $H_0 \ : \ \pi_1 = \pi_2$ vs $H_1 \ : \ \pi_1 > \pi_2$ (test unilatéral à droite).
En effet, dire que le vaccin 2 est plus efficace que le vaccin 1 revient à dire qu'après vaccination, la proportion π_2 de contaminés dans la population 2 est inférieure à la proportion π_1 de contaminés dans la population 1.

Nous pouvons raisonner en 4 étapes ou 6 étapes comme précédemment (bon exercice de rédaction). Pour gagner un peu de temps, utilisons l'onglet statistiques du software Geogebra.



La $p-$valeur i.e $P_{H_0 \text{ vraie}}(Z \geq 0,3286) = 0,3712 > \alpha = 0,05$. Donc au vu de nos échantillons et au risque α de 5%, on ne rejette pas H_0.
On ne peut pas conclure que le second vaccin soit plus efficace que le premier.

1.4.2.2 : Test de comparaison de deux variances $(H_0 : \sigma_1^2 = \sigma_2^2)$.

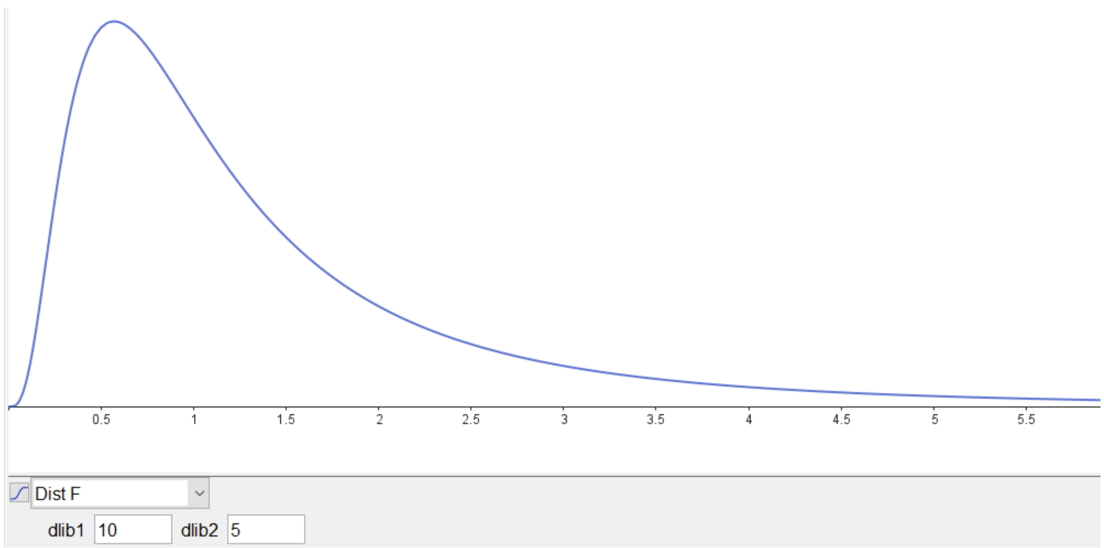
Nous cherchons à comparer les variances σ_1^2 et σ_2^2 de deux populations. Pour ce faire, nous prélevons deux échantillons indépendants de tailles n_1 et n_2 de chacune de ces populations.

Notons s_1^2 et s_2^2 (resp. $\hat{s_1}^2$ et $\hat{s_2}^2$) les variances (resp. les variances corrigées) des échantillons 1 et 2.

Théorème : Soit $\hat{S_1}^2$ (resp. $\hat{S_2}^2$) la variable aléatoire qui à tout EAS de taille n_1 de la population 1 (resp. à tout EAS de taille n_2 de la population 2) lui associe sa variance corrigée. Alors :

$$F = \frac{\hat{S_1}^2}{\hat{S_2}^2} \hookrightarrow \text{Fisher-Snedecor à } (n_1 - 1, n_2 - 1) \text{ ddl}$$

Nous mettrons en œuvre ce test dès le paragraphe suivant. Donnons cependant un aperçu de sa fonction de densité. Ici $F \hookrightarrow$ Fisher-Snedecor à $(10, 5)$ ddl



Remarque : Historiquement, on ne disposait que de tables statistiques de la loi F de Fisher-Snedecor, dont toutes les valeurs prises étaient supérieures ou égales à 1. On devait donc décider que la population 1 était celle dont l'échantillon avait la variance corrigée la plus grande. Auquel cas, la règle de décision était un peu plus simple, puisque le f_{obs} ne pouvait prendre que des valeurs supérieures à 1. Exercice : énoncez-la !

1.4.2.3 : Test de comparaison de deux moyennes $(H_0 : \mu_1 = \mu_2)$.

Les deux échantillons considérés par la suite sont supposés **indépendants** et issus de deux populations dont le caractère quantitatif étudié est **distribué normalement**.

Paramètre des 2 populations :
On note μ_1 (resp. μ_2) la valeur moyenne du caractère étudié dans la population 1 (resp. dans la population 2).
On note σ_1 (resp. σ_2) l'écart-type du caractère étudié dans la population 1 (resp. dans la population 2).

Paramètres des 2 échantillons :
On note $\bar{x}_1$ (resp. $\bar{x}_2$) la valeur moyenne du caractère étudié dans l'échantillon 1 (resp. dans l'échantillon 2).
On note $\hat{s}_1$ (resp. $\hat{s}_2$) l'écart-type corrigé du caractère étudié dans l'échantillon 1 (resp. dans

l'échantillon 2).
On note $\hat{\sigma}$ la *pooled-variance*, qui correspond à une variance pondérée et corrigée calculée à partir des tailles et des variances de chacun des 2 échantillons :

$$\hat{\sigma}^2 = \frac{SCE_{\text{totale}}}{\text{nombre de ddl}}$$

soit :

$$\hat{\sigma}^2 = \frac{(n_1-1)\hat{s}_1^2 + (n_2-1)\hat{s}_2^2}{(n_1-1)+(n_2-1)} = \frac{(n_1-1)\hat{s}_1^2 + (n_2-1)\hat{s}_2^2}{n_1+n_2-2}$$

ou en termes de variances non corrigées :

$$\hat{\sigma}^2 = \frac{n_1 s_1^2 + n_2 s_2^2}{n_1+n_2-2}$$

Variables de décision en fonction des cas :

σ_1 et σ_2 connus	H_1	Variable de décision sous H_0	$p-$valeurs
$Z-$test	$\mu_1 \neq \mu_2$	$Z = \dfrac{\bar{X}_1 - \bar{X}_2}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}}$	$P(\lvert Z\rvert > \lvert z_{obs}\rvert)$
	$\mu_1 > \mu_2$		$P(Z > z_{obs})$
	$\mu_1 < \mu_2$		$P(Z < z_{obs})$

σ_1 et σ_2 inconnus mais égaux	H_1	Variable de décision sous H_0	$p-$valeurs
$T-$test	$\mu_1 \neq \mu_2$	$T = \dfrac{\bar{X}_1 - \bar{X}_2}{\hat{\sigma}\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}}$	$P(\lvert T\rvert > \lvert t_{obs}\rvert)$
	$\mu_1 > \mu_2$		$P(T > t_{obs})$
	$\mu_1 < \mu_2$		$P(T < t_{obs})$

σ_1 et σ_2 inconnus et inégaux	H_1	Variable de décision sous H_0	$p-$valeurs
$T-$test de Welch	$\mu_1 \neq \mu_2$	$T = \dfrac{\bar{X}_1 - \bar{X}_2}{\sqrt{\dfrac{\hat{s_1}^2}{n_1} + \dfrac{\hat{s_2}^2}{n_2}}}$	$P(\lvert T\rvert > \lvert t_{obs}\rvert)$
	$\mu_1 > \mu_2$		$P(T > t_{obs})$
	$\mu_1 < \mu_2$		$P(T < t_{obs})$

Exemple 10 : nous souhaitons comparer les rendements moyens de blé tendre dans deux départements différents.

1. En Eure-et-Loir, sur 20 exploitations, nous avons obtenu un rendement moyen de 82 q/ha et un écart-type de 5 q/ha.
2. En Seine Maritime, sur 16 exploitations, nous avons obtenu un rendement moyen de 86 q/ha et un écart-type de 8 q/ha.

Peut-on considérer au risque de première espèce de 5% que les rendements de blé tendre en Seine maritime sont supérieurs à ceux d'Eure-et-Loir ?

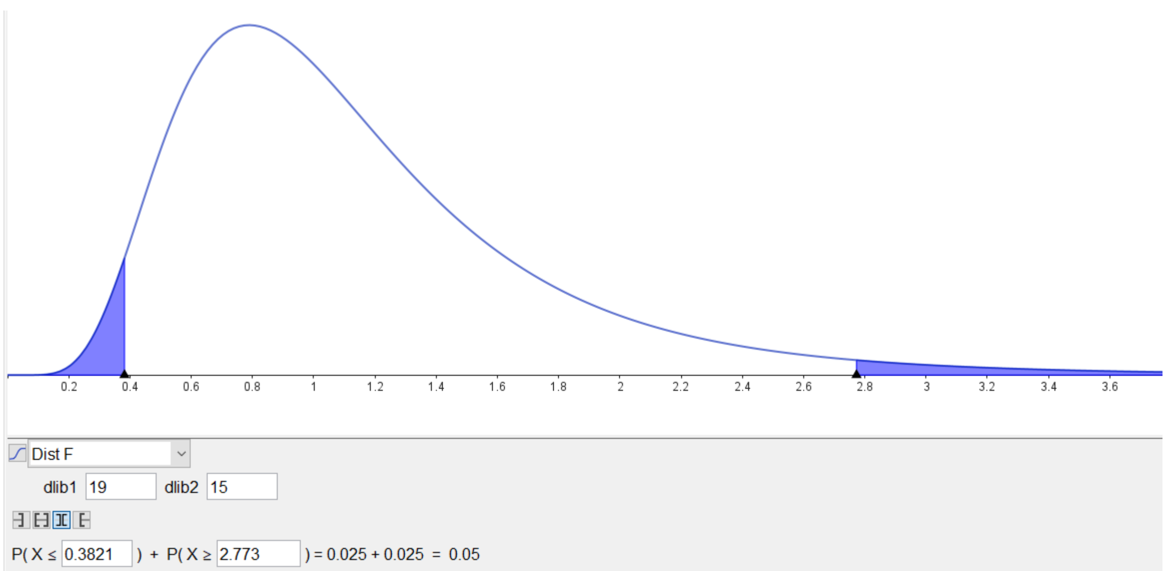
Solution : Appelons population 1 (resp. population 2) l'ensemble des exploitations produisant du blé tendre en Eure-et-Loir (resp. en Seine maritime).
Remarquons que **nous ne connaissons rien des deux populations** : σ_1 et σ_2 sont inconnues (tout comme μ_1 et μ_2) !
En revanche, nous connaissons tout des échantillons :

— $n_1 = 20$, $\bar{x_1} = 82$ q/ha et $s_1 = 5$ q/ha,
— $n_2 = 16$, $\bar{x_2} = 86$ q/ha et $s_2 = 8$ q/ha.

Étape 1 : Test d'égalité des variances.

Soient σ_1^2 et σ_2^2 les variances des productions de blé tendre en Eure-et-Loir et en Seine maritime.

1. $H_0 \; : \; \sigma_1^2 = \sigma_2^2$ vs $H_1 \; : \; \sigma_1^2 \neq \sigma_2^2$.
2. on se fixe un risque de première espèce $\alpha = 5\%$.
3. Soient $\hat{S_1}^2$ (resp. $\hat{S_2}^2$) la variable aléatoire qui à tout EAS d'effectif 20 de la population 1 (resp. d'effectif 16 de la population 2), lui associe sa variance corrigée.
   Sous H_0 : $F = \dfrac{\hat{S_1}^2}{\hat{S_2}^2} \hookrightarrow$ Fisher-Snedecor à $(19, 15)$ ddl
4. On détermine les zones de rejet et de non-rejet de H_0 à l'aide d'un software.



— Si $f_{obs} \in [0,3821; 2,773]$, on ne rejette pas H_0,
— Si $f_{obs} \notin [0,3821; 2,773]$, on rejette H_0.

5. Pour calculer f_{obs}, nous avons besoin de calculer $\hat{s_1}^2 = \dfrac{20}{19} s_1^2$ et $\hat{s_2}^2 = \dfrac{16}{15} s_1^2$. On trouve $\hat{s_1}^2 = 26,31$ et $\hat{s_2}^2 = 68,27$. Ainsi, $f_{obs} = \dfrac{26,31}{68,27} = 0,385$.
6. $f_{obs} = 0,385 \in [0,3821; 2,773]$, donc au risque de première espèce de 5% et au vu de nos deux échantillons, on ne rejette pas H_0. Nous pouvons donc supposer que les variances σ_1^2 et σ_2^2 sont égales.

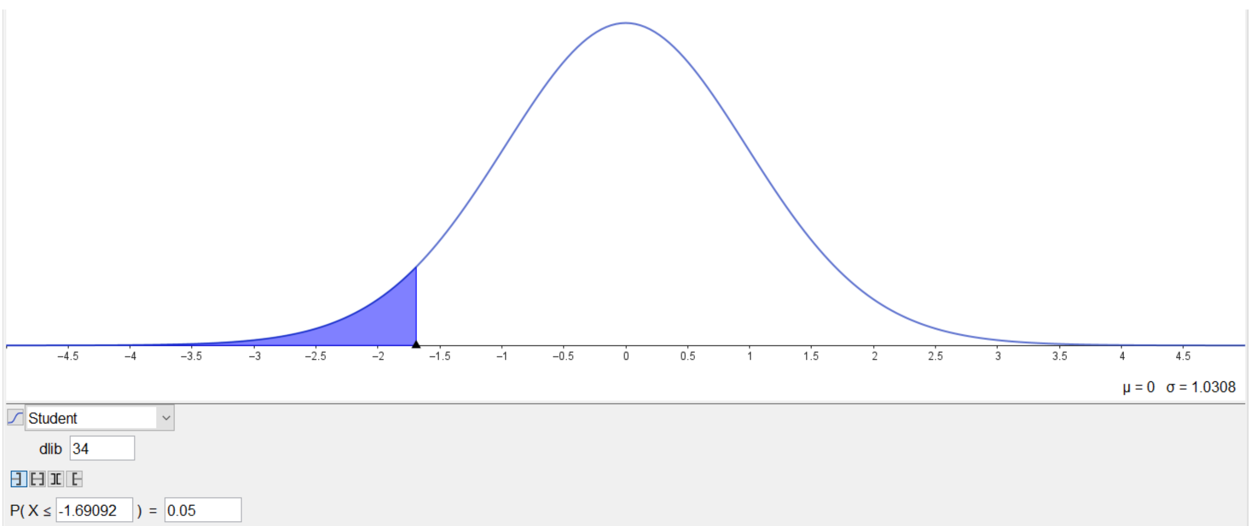
Comme précisé auparavant, nous n'avons pas démontré que les variances σ_1^2 et σ_2^2 sont égales, mais nous ne pouvons pas dire le contraire, même si f_{obs} n'était pas loin d'une des deux valeurs critiques.

Étape 2 : Test de comparaison des moyennes.

Soient μ_1 et μ_2 les rendements moyens de blé tendre en Eure-et-Loir et en Seine maritime. Au vu de l'étape 1, nous allons utiliser le T-test de comparaison des moyennes.

Calcul préliminaire de la pooled-variance : $\hat{\sigma}^2 = \dfrac{20 \times 5^2 + 16 \times 8^2}{20 + 16 - 2} \approx 44,82$. D'où $\hat{\sigma} \approx 6,7$.

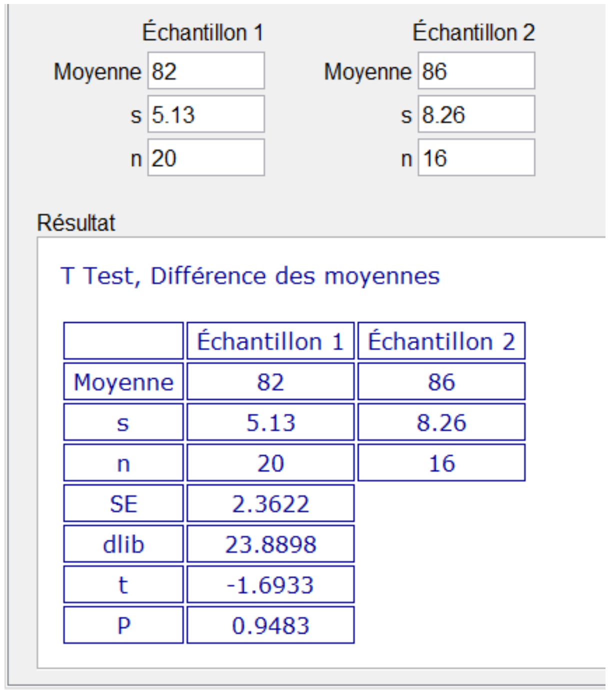
1. $H_0 : \mu_1 = \mu_2$ vs $H_1 : \mu_1 < \mu_2$
2. on se fixe un risque $\alpha = 5\%$
3. Soit $\bar{X}_1$ (resp. $\bar{X}_2$) la variable aléatoire qui à tout EAS de 20 exploitations d'Eure-et-Loir (resp. de 16 exploitations de Seine maritime) associe le rendement moyen de blé tendre.
   Sous H_0 : $T = \dfrac{\bar{X}_1 - \bar{X}_2}{\hat{\sigma}\sqrt{\dfrac{1}{20} + \dfrac{1}{16}}} \hookrightarrow$ Student à 20+16-2 = 34 ddl
4. On détermine les zones de rejet et de non-rejet de H_0 à l'aide d'une table statistique ou d'un software :



   — Si $t_{obs} \geq -1,69$, on ne rejette pas H_0,
   — Si $t_{obs} < -1,69$, on rejette H_0.
5. $t_{obs} = \dfrac{82 - 86}{6,7 \times \sqrt{\dfrac{1}{20} + \dfrac{1}{16}}} \approx -1,78$.
6. $t_{obs} = -1,78 < -1,69$, donc au risque de première espèce de 5% et au vu de nos échantillons, on rejette H_0. Nous pouvons dire que le rendement moyen de blé tendre en Seine maritime est supérieur à celui à Eure-et-Loir.

Remarque importante : on peut à juste titre objecter le fait qu'on ait pris pour hypothèse l'égalité des variances de chaque population, puisque conserver $H_0 : \sigma_1^2 = \sigma_2^2$ ne prouve rien en fait sur sa véracité. Auquel cas, on peut directement appliquer le test de Welch.

En utilisant l'onglet statistiques de Geogebra, avec $n_1 = 20$, $n_2 = 16$, $\bar{x_1} = 82$, $\bar{x_2} = 86$, $\hat{s_1} = \sqrt{26,31} = 5,13$ et $\hat{s_2} = \sqrt{68,27} = 8,26$, nous obtenons :



	Échantillon 1	Échantillon 2
Moyenne	82	86
s	5.13	8.26
n	20	16
SE	2.3622	
dlib	23.8898	
t	-1.6933	
P	0.9483	

Et là, ô (petite) surprise, en utilisant ce test, nous trouvons une p-valeur de 0,9483 largement supérieure au risque α de 5%. Bref, nous ne pouvons pas rejeter H_0 et dire que le rendement moyen de blé tendre est supérieur en Seine maritime !
Nous reviendrons sur cette contradiction apparente en TP.

2 Corrigé des quiz et compléments

2.1 Corrigés des quiz et exercices

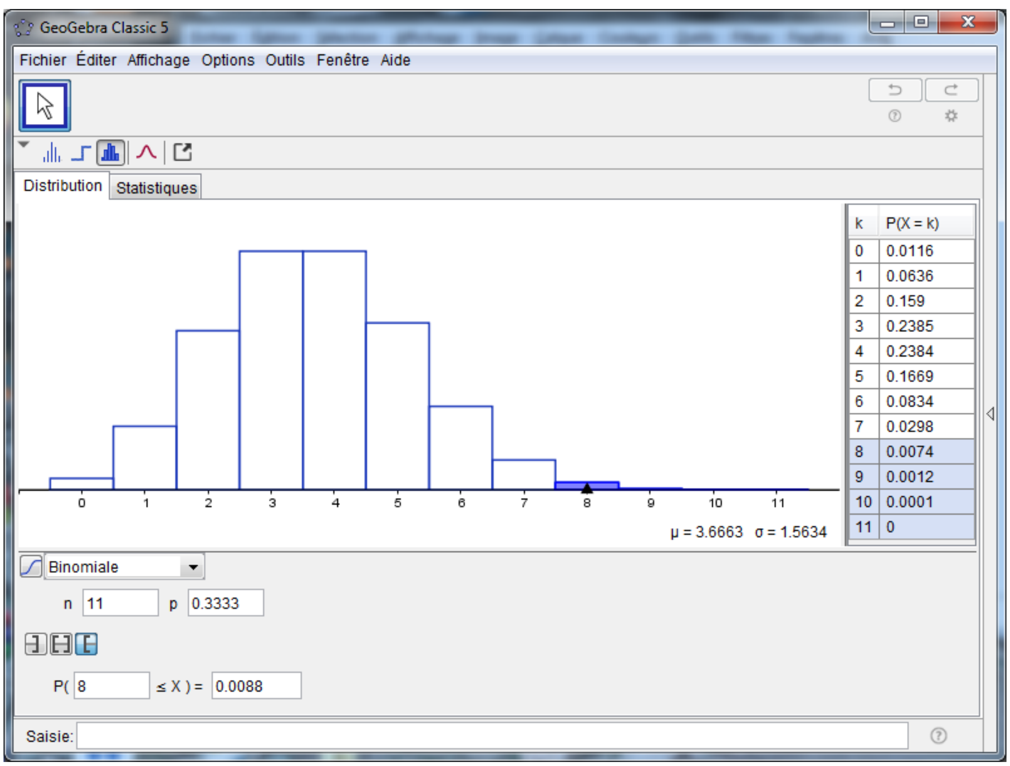
Quiz 1 : corrigé rapide.

1. (c)
2. (b)
3. (c) : $0,005 \times 0,997 \approx 0,00499$
4. (b) : $0,9552/(1-0,005) = 0,96$
5. (a) (d) : $0,995 \times 0,96 + 0,005 \times 0,003 = 0,9552$
   (b) (b) : $1 - 0,9552^{10} = 0,3677$

Quiz 2 : Les calculs ont été effectués à l'aide du software Geogebra.

1. (d) : 0,0028
2. (b) : 0,9869
3. (c) : 1,107 mm
4. Augmenter la taille n de l'échantillon : (b) diminue le risque α et (d) diminue le risque β (donc augmente la puissance du test).
5. (a) P est croissante

Exercice 1 : on évalue $P(X \geq 8)$ où $X \hookrightarrow \mathcal{B}(11; 1/3)$.



$P(X \geq 8) = 0,0088 < \alpha = 0,01$ et $P(X \geq 7) = 0,0386 > 0,01$. CQFD.

Quiz 3 : Les calculs ont été effectués avec le software Geogebra

1. (c) $28/200 = 0{,}14$
2. (b) $\pi > 0,11$ est la plus raisonnable. On peut aussi poser (d) $\pi \neq 0,11$
3. (d) $P(F \geq 0,14) = P(X \geq 0,14 \times 200) = P(X \geq 28) = 0,1094$, où $X \hookrightarrow \mathcal{B}(200; 0,11)$ (loi exacte)
4. (b) Faux. On rejette H_0 si $f \geq 30/200 = 0,15$. Or $f = 0,14 < 0,15$
5. (d) (0,11 ; 0,022) : $E(F) = \pi$ et $\sigma(F) = \sqrt{\pi(1-\pi)/200}$
6. (a) $]1,645; +\infty[$: 1,645 est le fractile d'ordre 0,95 de U
7. (a) Oui, même conclusion qu'avant. $u_{obs} = (0,14 - 0,11)/0,022 = 1,36 < 1,645$

Quiz 4 : Les calculs ont été effectués avec le software Geogebra

1. (c) C'est un test unilatéral à gauche
2. (d) $E(\bar{X}) = 2,7$ et $\sigma(\bar{X}) = \dfrac{\sigma}{\sqrt{20}} = 0,056.$
3. (a) 0,0101
4. (a) $-2,3255$
5. (a) Vrai : par translation et homothétie de rapport positif
6. (b) Vrai
7. (b) $[-1,96; 1,96]$
8. (a) et (b)

Quiz 5 : Les calculs ont été effectués avec le software **R**

1. (c) Sous H_0, les données obtenues sont improbables (seuil α fixé à l'avance). La notion de $p-$valeur ne porte pas sur H_1 donc on ne retient pas (d). Il est ainsi classique, quand cela est possible de calculer la puissance du test après le rejet de H_0.
2. (b) : "sépare" bien les courbes sous $H_0 : \mu = 2,5$ et $H_1 : \mu = M$ et (d) : "rend pointues" les gaussiennes.
3. (a) distribution uniforme (plate)
4. (a) Vrai
5. (a) Vrai

Quiz 6 : Un florilège d'idées fausses !

1. (b) Faux ! Avoir une $p-$valeur de 0,03 signifie que <u>sous H_0</u>, la probabilité d'obtenir nos données ou plus extrêmes est de 0,03.
2. (b) Faux ! Il est possible d'obtenir une grande $p-$valeur même si H_0 est fausse. Une des causes possibles est le manque de puissance de notre test : *c.f* graphiques de distribution des $p-$valeurs.
3. (b) Faux !
4. (a) Vrai
5. (a) Vrai

Exercice 2 : Démarche standard

1. $\bar{x} = 249,64g$ et $s_x = 1,394g$ (l'écart-type corrigé est $\hat{s}_x = 1,447g$)
2. On suppose ici $\sigma = 2g$. Nous testons $H_0 : \mu = 250$ contre $H_1 : \mu < 250$.
   Comme l'écart-type de la population est connu, la variable de décision sous H_0 est $U = \dfrac{\bar{X} - \mu}{\sigma/\sqrt{14}} \hookrightarrow \mathcal{N}(0;1)$.
   Si $u_{obs} < -1,645$, on rejette H_0, sinon on ne rejette pas H_0. $u_{obs} = (249,64 - 250)\sqrt{14}/2 = -0,673 \geq -1,645$, donc au risque de première espèce de 5% on ne rejette pas H_0. On ne peut pas suspecter le boulanger de fraude.
3. Cette fois-ci, l'écart-type de la population est inconnu. La variable de décision est $T = \dfrac{\bar{X} - 250}{\hat{S}/\sqrt{14}} \hookrightarrow$ Student à 14-1 = 13 ddl.
   En utilisant la table de la loi de Student ou un software, on obtient la règle de décision suivante : si $t_{obs} < -1,77$, on rejette H_0, sinon on ne rejette pas H_0. $t_{obs} = (249,64 - 250)\sqrt{14}/1,447 = -0,93 \geq -1,77$, donc au risque de première espèce de 5% on ne rejette pas H_0. La conclusion est la même que précédemment.

Quiz 7 : Se servir d'un intervalle de confiance pour un test bilatéral !

1. (d) 0,0717
2. (b) $\chi^2_{15;0,975} = 27,488$, d'où $16 \times 0,0717/27,488 = 0,041$ par défaut.
3. (a) Oui. La borne supérieure de l'intervalle de confiance de σ^2 au seuil de confiance de 95% est 0,184 par excès. $0,204 \notin [0,041; 0,184]$, donc on rejette H_0. La démarche classique en 4 ou 6 étapes aurait abouti à la même conclusion : $k_{obs} = 16 \times 0,0717/0,204 = 5,623 \notin [6,262; 27,488]$.

Exercice 3 : Encore les intervalles de confiance.

1. $\hat{s}^2 = 3,0186$.
2. $IC_{0,95} = [1,82; 5,94]$
3. $2,5^2 = 6,25 \notin [1,82; 5,94]$, donc au risque de première espèce de 5%, on rejette $H_0 : \sigma^2 = 2,5^2$.

Quiz 8 : Toute proportion gardée ...

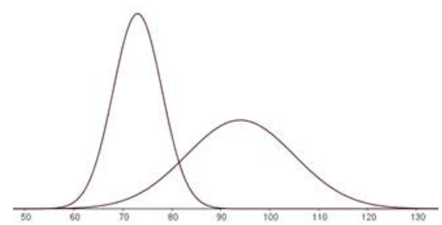
1. (b) 14/150 = 0,0933
2. (a) 0,1324
3. (c) unilatérale à gauche. On cherche à savoir si $\pi < 0,11$.
4. (a) $[\![0;9]\!]$: $P(X \leq 9) < 0,05$ et $P(X \leq 10) > 0,05$, où $X \hookrightarrow \mathcal{B}(150; 0,11)$.
5. (b) Faux : 14 > 9.
6. (d) $u_{obs} = \dfrac{0,0933 - 0,11}{\sqrt{0,0933(1-0,0933)/150}} = -0,6537$
7. (b) Faux : $u_{obs} \in]-1,645; +\infty[$ (zone de non-rejet de H_0).

Quelques exercices non corrigés ...

Règle de décision : La taille d'un crapaud commun (Bufo bufo L.) est la mesure en millimètre de la distance de l'extrémité du museau à la pointe de l'urostyle de l'animal. Les naturalistes ont mis en évidence qu'à l'âge adulte :

- la taille des femelles est distribuée selon une loi normale d'espérance 94 et d'écart-type 11,
- la taille des mâles est distribuée selon une loi normale d'espérance 73 et d'écart-type 5.

1. Écrire une légende pour le graphique ci-contre :



2. (a) Donner le pourcentage de femelles ayant une taille comprise entre 90 et 120 millimètres,
   (b) Donner le pourcentage de mâles ayant une taille comprise entre 70 et 80 millimètres,
   (c) Déterminer le premier et le neuvième décile de la distribution des tailles des crapauds mâles. *On rappelle que le premier décile d'une distribution est la valeur en dessous de laquelle on trouve 10% des tailles et que le neuvième décile est celle en dessous de laquelle on trouve 90% des tailles.*
3. Pour sexer rapidement un crapaud adulte, un naturaliste adopte la règle de décision suivante :
   Si la taille du crapaud est inférieure à 80 mm alors il admet que c'est un mâle.
   Sinon il admet que c'est une femelle.
   Représenter graphiquement et déterminer les probabilités des évènements suivants :
   (a) Un crapaud mâle pris au hasard a une taille supérieure à 80 mm
   (b) Un crapaud femelle pris au hasard a une taille inférieure à 80 mm
4. En supposant que le sex ratio des crapauds communs est de sept mâles pour une femelle, déterminer la probabilité de se tromper lorsque l'on applique cette méthode. *On pourra s'aider d'un arbre pondéré.*

Interprétation : On donne ci-dessous plusieurs informations. Entourez celles qui sont vraies.

1. Affirmation 1 : Ne pas rejeter $H_0 : \mu = \mu_0$, c'est prouver qu'il n'y a pas de différence significative entre μ et μ_0.
2. Affirmation 2 : Rejeter H_0, c'est prouver que notre échantillon était "exceptionnel".
3. Affirmation 3 : Dans une ville donnée, la proportion de gauchers est $p = 0,2$. On prélève un échantillon aléatoire simple de 200 individus de cette ville. Un intervalle de fluctuation asymptotique au seuil de confiance de 95% de p est [0,1445 ; 0,2555].
4. Affirmation 4 : Dans une autre ville, on prélève un échantillon aléatoire simple de 80 individus. On s'aperçoit qu'il contient 14 gauchers. Un intervalle de confiance au seuil de confiance de 95% de p, la proportion de gauchers dans la ville, est [0,0969 ; 0,2531]
5. Affirmation 5 : Plus la puissance d'un test est grande, plus on rejette H_0 à raison.

Puissance et taille : Un éleveur breton utilise une ration A pour engraisser ses porcs. Il estime l'écart-type de leurs masses à 20 kg. Il veut tester une ration B avec laquelle il espère un gain de masse de 12 kg. Quelle doit-être la taille de l'échantillon de son expérimentation afin d'avoir une puissance d'au moins 0,8 ?

Tests de conformité :

1. Nous prélevons 150 bouteilles dans une production de bouteilles d'eau dont on mesure le pH. Nous observons que 25 bouteilles ont un pH supérieur à 7,5. Peut-on dire, au risque de première espèce de 5% que la proportion de bouteilles dont le pH dépasse 7,5 est supérieure à 10% ?
2. Afin de vérifier la qualité du lait d'une exploitation agricole, on effectue des prélèvements journaliers pour lesquels on dose l'acidité du lait.
   On considère que le lait est « suspecté de mouillage » (dilution frauduleuse par l'agriculteur) si son dosage moyen d'acidité est inférieur à 12 degrés Dornic.
   Voici les résultats obtenus sur 15 prélèvements : 13 15 11,5 13 10 9,5 10 10 12 11,5 12,5 14 11 10 12
   On supposera que le dosage de l'acidité du lait est une variable aléatoire normalement distribuée.
   Peut-on au risque de première espèce de 5% suspecter le lait de cette exploitation de mouillage ?

Tests de comparaison :

1. Une firme pharmaceutique met au point un somnifère qu'elle teste sur deux groupes de patients, issus de deux populations :
   L'échantillon 1, issu de la population 1, contient 150 patients et à l'issue du test, 110 d'entre eux ont vu leur sommeil réparé de manière significative.
   L'échantillon 2, issu de la population 2, contient 280 patients et à l'issue du test, 200 d'entre eux ont vu leur sommeil réparé de manière significative.
   Au vu de cet échantillon, et au risque de 5%, peut-on affirmer que la population 2 est moins sensible au somnifère que la population 1 ?
2. On s'intéresse aux rendements de Cabernet Franc dans 10 exploitations du Sud-Ouest et de 16 exploitations du Languedoc.
   On a obtenu les résultats suivants :
   Pour les exploitations du Sud-Ouest, le rendement moyen a été en 2020 de 56,2 hl/ha avec un écart-type de 2,5 hl/ha.
   Pour les exploitations du Languedoc, le rendement moyen a été en 2020 de 53,5 hl/ha avec un écart-type de 1,5 hl/ha.
   Peut-on affirmer, avec un risque de première espèce de 5 % que le rendement de Cabernet Franc dans le Sud-Ouest est supérieur à celui du Languedoc ?

2.2 Bibliographie

1. S. Baggio / S. Deline / S. Rothen - SUP en poche L1/L2 - Statistique inférentielle - deboeck supérieur (2023)
2. F. Bertrand / M. Maumy-Bertrand - MAXI FICHES - Statistique pour les scientifiques - Dunod (2011)
3. J-C Fauré - TEAM - Mathématiques BTSA - 3. Tests statistiques - TEC & DOC (2001)
4. J. Pagès - Statistique générale pour Utilisateurs - 1. Méthodologie - PUR (2010)